The Effect of Energy Input on the Earth's Upper Atmosphere

by

Jie Zhu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Atmospheric, Oceanic and Space Sciences)
in The University of Michigan
2016

Doctoral Committee:

Professor Aaron J. Ridley, Chair
Associate Professor John E. Foster
Professor Mark B. Moldwin
Associate Professor Shasha Zou

ACKNOWLEDGEMENTS

There have been so many people I would like to extend my thanks to, who have been generously giving me help through the completion of this dissertation.

I would like to express my sincerest gratitude to my advisor, Professor Aaron J. Ridley for his tremendous guidance, encouragement and support throughout my graduate study. He is one of the most knowledgable and insightful persons I have ever met. Every time when I got stuck in my research, discussion with him always helped me come out with efficient and feasible solutions. He has been not only an excellent academic advisor, but also a great mentor. He saw my talents, provided me lots of opportunities to present findings, and also offered candid advice on where I need to improve to be a good scientist. He is very humorous and helpful. Working with him has been always an enjoyable and cheerful thing.

A great deal of thanks go to the rest of my dissertation committee members, Professors Mark B. Moldwin, Shasha Zou and John E. Foster, for their time and efforts on improving my thesis by providing insightful comments and suggestions. I would like to especially thank Professor Zou, who has been giving me lots of insights on substorm and has encouraged me in many ways. I appreciate Professor Moldwin's intriguing lectures on space environment and also his many support on my internship and job applications.

I would also like to thank Andrew Nagy, Darren DeZeeuw, Mike Liemohn, Gabor Goth for their kind help on resolving my scientific and technical issues. Professor Nagy gave me great amounts of help on searching for a better formula of electron

photoelectron heating efficiency when I was developing the ionosphere temperature model. I also appreciate many faculties in the department, from whom I have received many excellent lectures and instructions on space science.

Many thanks to Laura Hopkins, Sandra Pytlinski, Sandee Hicks, Rachel Long, Eidilia Thomas and Jan Beltran for their administrative support, and to Darren Britten-Bozzone, Faye Ogasawara and other IT specialists for their timely help with my computer issues. Without them, I could not have conducted my research so successfully.

I would like to say thank you to many of my friends who have been supporting me from every aspect in graduate school, including Xiangyun Zhang, Shaosui Xu, Nicholas Perlongo, Charles Bussy-Virat, Casey Steuer, Nathan Boll, Yuxi Chen, Xing Meng, Fang Pan, Zhengguang Huang, Yinsi Shou, Chuangfei Dong and Jin Meng, to name a few.

Finally, I would like to express my deep gratitude to my family, my mother Meidi Wang, my father Yusong Zhu and my husband Xiaojian Liu, for their constant support and encouragement keeping me going through all of the difficult periods during my graduate study.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

The Effect of Energy Input to the Earth's Upper Atmosphere

by

Jie Zhu

Chair: Aaron Ridley

The Earth's upper atmosphere is affected by a variety of energy inputs, mainly incident solar irradiance, dissipation of internal gravity waves, solar wind electric fields and energetic particle precipitation. This dissertation consists of four studies focusing on the response of the coupled thermosphere ionosphere system to several energy inputs by using the Global Ionosphere Thermosphere Model (GITM).

Two studies focus on how a sudden increase in solar EUV emissions affects the ionospheric density and thermospheric dynamics during a solar flare. Both the flare incident energy and the background conditions, i.e., the geomagnetic fields and seasons, control the ionospheric response. Solar flares with different incident energies into the same background hemisphere caused ionospheric disturbances of similar profiles but different magnitudes, while the same flare spectra caused perturbations of similar magnitude but different profiles in different backgrounds. In addition, large scale gravity waves are shown to travel from the dayside to the nightside when a flare occurs. Acoustic waves are modulated by neutral-ion momentum coupling because ion drag can be comparable to the gradient in pressure in the acoustic temporal scale.

Ionospheric temperatures are strongly coupled with the neutral atmosphere and

control many chemical and physical processes. An improved ionospheric energy model was developed in GITM and is described here. The model allows self-consistent energy coupling between electrons, ions and neutrals. The model was applied to explore the ionospheric response to an idealized substorm. The ion temperature was controlled mainly by the solar wind electric field, while the electron temperature was controlled by particle precipitation. In addition, the performance of a commonly used neutral-ion collisional heating approximation was investigated. The global averaged neutral temperature could be underestimated by approximately 10% due to the neglect of the indirect heating from electrons through ions. The use of a photoelectron heating efficiency for neutrals by global IT models that use Joule heating for energy coupling between the thermosphere and ionosphere, was essentially a compensation for the neglect of this heating.

CHAPTER I

Introduction

The earliest reliable account of ionospheric phenomenon was the the aurora borealis, dating back to around 568 BC (*Stephenson et al.*, 2004). Auroras are displays of natural light appearing in the high-latitude night sky, which result from ionization and excitation of the neutral atmosphere by energetic particle precipitation when the near Earth's environment is sufficiently disturbed by solar wind. The discovery of the Earth's ionosphere started in 1899 when Gulglielmo Marconi achieved the first trans-Atlantic communication using radio waves (*Marconi*, 1909). It was soon realized by the science community that there exists a permanent electrically conducting layer high in the atmosphere, which is produced by the upper atmosphere photoionized by extreme ultraviolet light from the sun. This conducting layer, named the "Ionosphere" by R. A. Watson-Watt in 1926 (*Gardiner*, 1969), deflects radio waves around the Earth at great spatial scales. Sydney Chapman derived the first fundamental mathematical model of the ionosphere in 1931 (*Chapman*, 1931). With the development of rocket and satellite technologies, as well as ground-based instruments, our knowledge about the upper atmosphere has greatly increased. Variations and multi-scale features of the thermosphere and ionosphere were discovered, such as diurnal and seasonal variations, Sq currents and high-latitude convection patterns. Empirical models of the ionosphere and thermosphere have been developed through

statistical analysis of extensive measurement datasets since the 1970s (*Hedin et al.*, 1974; *Hedin*, 1983; *Bilitza*, 1986). Some of the statistical models describe monthly average conditions based on solar activities and geomagnetic disturbance indexes, which provide scientists and engineers easy access to climatological conditions of the upper atmosphere. Major advances in high-performance computers enable the development of high-resolution physical models of the thermosphere and ionosphere based on first principles since the 1980s. Physical models implement basic chemistries, dynamic equations and multiple drivers from the atmosphere below and from the magnetosphere above.

In this chapter, the general structures and some important dynamics of the thermosphere and ionosphere will be discussed.

1.1 Structure of upper atmosphere

The Earth's atmosphere is a very thin layer of air (typically less than 1000 km) that surrounds our planet. The two major gases making up 99% of the volume of air of the Earth's atmosphere are nitrogen ($\sim 78\%$) and oxygen ($\sim 21\%$). The other 1% is made up by minor gases such as water vapor, carbon dioxide and ozone. Based on the vertical variations in neutral temperature, the atmosphere is divided into four layers with respect to altitude: troposphere, stratosphere, mesosphere and thermosphere. Figure 1.1 shows the vertical structure of the global mean temperature of the Earth's atmosphere.

The troposphere is the lowest layer, which extends from the Earth's surface to about 15 km altitude. In this region, the temperature gradually decreases with increasing height from 290 K on the Earth's surface to 210 K at the tropopause. The main energy source in this region is the Earth's surface. The majority of the solar radiation is absorbed by the Earth's surface, but only a small portion is directly absorbed by the neutrals when first entering into the atmosphere. This makes the

Earth's surface a heat source for the atmosphere on the dayside. The heat is transferred from the Earth's surface to the troposphere through convection, conduction and infrared radiation, which results in a warmer troposphere near the Earth's surface and cooler towards the tropopause. The troposphere also loses energy through infrared radiation and evaporation of H_2O.

Above the troposphere is the stratosphere where the temperature steadily climbs. The increase of temperature in this layer is caused by the strong absorption of solar ultraviolet radiation by ozone. At the stratopause, the ozone heating reaches a maximum, leading to a temperature peak at about 45 km. The thickness of this layer is about 30 km.

Above the stratosphere is the mesosphere. The mesosphere temperature decreases as the altitude increases. In this region, there are three major energy sources: solar UV absorption by ozone at lower heights, heat conduction down from the upper thermosphere, and chemical heating. The upper boundary of the mesosphere, the mesopause, is the coldest region of the Earth's atmosphere with temperatures of 150-200 K.

1.1.1 Thermosphere

The lower boundary of the thermosphere starts from the mesopause. The thermospheric temperature increases with altitude. The maximum temperature can range roughly from 600 K to 1200 K from solar minimum to solar maximum. The thermosphere absorbs solar EUV (200-1000 $\AA$) and UV (1200-2000 $\AA$) leading to photoionizations and photodissociations. The thermosphere is also a sink of magnetospheric energy including Joule heating by auroral electrical currents and particle precipitation from the magnetosphere at high latitudes (*Killeen et al.*, 1984). Tides, planetary waves and gravity waves propagate from the lower atmosphere and dissipate energy in the thermosphere. Molecular heat conduction into the lower thermosphere and

mesosphere is the main cooling process of the upper thermosphere. Radiative cooling due to minor constituents such as NO and CO_2 plays an important role in the thermal budget of the thermosphere, especially during periods of high solar activity or geomagnetic disturbances (*Meada*, 1992).

In the thermosphere, the air density is large enough to be considered a collisional medium. The thermosphere transition to the exosphere as the mean free path reaches one pressure scale height. The upper boundary of the thermosphere varies according to solar activity. When the sun is around the peak of the sunspot cycle, the thermopause can be raised to heights around 1000 km. When the sun is least active, the thermopause declines to about 500 km. The exosphere is the region where particles may escape from the Earth's atmosphere. This is caused by the kinetic energy of particles exceeding the gravitational binding energy, and the length of the mean free path is sufficiently long to avoid collisions with other particles.

The thermosphere is usually in a hydrostatic equilibrium state, which means that the force of the gravity is balanced by the force of the pressure gradient in the vertical direction. Assuming the direction of the increasing altitude as positive, the vertical momentum equation can be written as:

$$\frac{dP}{dz} = -\rho g, \tag{1.1}$$

where P is the pressure, z is the altitude, ρ is the mass density and g is the acceleration due to gravity. Assuming the ideal gas law holds, the pressure can be expressed as:

$$P = n\kappa T, \tag{1.2}$$

where n is the number density, κ is the Boltzman's constant and T is the temperature. Equation 1.1 can be further written as:

$$\frac{1}{P}\frac{dP}{dz} = -\frac{1}{H}, \tag{1.3}$$

where H is the scale height:

$$H = \frac{\kappa T}{mg}. \tag{1.4}$$

Integrating P with altitude results in pressure exponentially decreasing with height, i.e.,

$$P = P_0 e^{-\frac{z-z_0}{H}}, \tag{1.5}$$

where the subscript zero refers to a reference height. This is the hydrostatic law or barometric law (*Kundu and Cohen*, 2008). Similarly,

$$n = n_0 \frac{T_0}{T} e^{-\frac{z-z_0}{H}}. \tag{1.6}$$

for an isothermal thermosphere,

$$n = n_0 e^{-\frac{z-z_0}{H}}, \rho = \rho_0 e^{-\frac{z-z_0}{H}}. \tag{1.7}$$

Below the homopause, the composition of the atmosphere remains unchanged by Eddy diffusion (or turbulent diffusion). In other words, the decreasing rate of individual constituent with height is the same for all species. In the thermosphere, individual constituents tend to distribute vertically according to its own scale height. Figure 1.2 shows the composition and the density profiles in the upper atmosphere. Lighter constituents, with larger scale heights, have the tendency to decrease slower with altitude than heavier compositions. The scale height is highly dependent on the solar activity level, since it is dependent on the temperature. The thermosphere tends to be more expanded due to greater scale heights during periods of higher solar activity.

1.1.2 Ionosphere

1.1.2.1 Basic profile

Within the upper mesosphere and thermosphere, a layer characterized by a high density of free electrons and free ions forms due to energetic photoionizations by solar radiance in UV and X-rays wavelengths. The ionosphere roughly extends from 90 km to 600 km. The reverse process of photoionization is recombination, in which neutral particles are produced by free electrons combining with positive ions. There are two ways of recombination: radiative recombination and dissociative recombination. Radiative recombination occurs when the electrons directly interact with the positive ions to produce neutral atoms or molecules and releases the excess energy in the form of light. The radiative recombination rates are relatively small and are negligible in most cases. Dissociative recombination is the dominant electron and ion loss process in the ionosphere. It involves free electrons combining with positive molecules to produce two neutral atoms. The dissociative recombination rates are approximately five orders of magnitude greater than the radiative recombination rates.

Figure 1.3 shows typical mid-latitude electron density profiles during periods of moderate solar activity. There are five primary ionospheric regions: D region (∼60-90 km, peaks around 90 km), E region (∼90-140 km, peaks around 110 km), F_1 region (140-200 km, peaks around 160 km), F_2 region (∼200-500 km, peaks around 300 km) and topside ionosphere above the F_2 region. The nightside ionosphere is generally smaller than the dayside ionosphere depending on altitude and latitude. The solar spectrums in X-ray (0.1 nm to 17 nm) and EUV (17 nm to 175 nm) wavelengths are the main source of ionizing neutral species in the upper atmosphere. Figure 1.4 shows the typical composition of the dayside ionosphere during solar minimum (*Johnson*, 1966). In the D region, the major positive ions are O_2^+ and NO^+. Hydrated positive ions dominate the lower D region. The main ionization sources of the D region are

X-rays, Lyman-α, precipitating electrons from the magnetosphere and galactic cosmic rays.

The E region is formed by solar X-rays, EUV (Lyman-β) and energetic particle precipitation in the auroral zone. The dayside E region (excluding the auroral oval) is a typical Chapman layer where the altitude distribution of ionization produced solely by the absorption of solar radiation (*Chapman*, 1931). Some of the basic assumptions used to derive the model are that the solar radiation is monochromatic and parallel, the neutral atmosphere is horizontally stratified and isotherm (i.e., distributed with a constant scale height). The dominant ion species in this region are O_2^+ and NO^+. The chemical reactions in the E region are very rapid: the chemical time scales are seconds. The E region largely weakens during the night due to the lack of an ionization source at low and middle latitudes. The E region in the auroral oval is strongly affected by particle precipitaion.

The F_1 region is also a Chapman layer. The major source for ionization in this region is solar EUV at 17-80 nm. NO^+ is the dominant ion in the lower F_1 regions, while above about 180 km, O^+ rapidly becomes dominant due to its increased lifetime. The F_2 layer is formed by the same radiation as the F_1 layer. The major ion in the F_2 layer is O^+, which contributes to the majority of the F_2 peak between 200 km to 400 km. The main loss of O^+ is through charge exchange with neutral molecular species, which fall off rapidly with altitude. The molecular ions then rapidly dissociatively recombine with electrons. This means that any vertical shift in the ionosphere can change the O^+ density significantly. At mid-latitudes, the height of the F_2 peak can be shifted by neutral winds. A meridional wind blowing equatorward (pole-ward) pushes the ionosphere up (down) along the magnetic field lines away from (toward) neutral molecular species, which raises (decreases) the density of the F_2 peak due to slower (faster) charge exchange rates at higher (lower) altitudes. Electric field is another source shifting the F_2 peak: an eastward (westward) electric field causes

upward (downward) ion **ExB** drift around the magnetic equator and at low-latitudes, leading to a vertical shift in the ionosphere, thus a density change of the F_2 peak.

1.1.2.2 Ground-based measurements

Basic parameters of the ionosphere can be measured by an incoherent scatter radar (ISR). ISRs detect the incoherent backscatter from free electrons in the Earth's ionosphere. A radar transmits a certain frequency of radio wave into the ionosphere, and receive reflections from free electrons in the ionosphere. The received echo is spread in frequency by the Doppler shifts associated with the thermal motion of electrons (*Gordon*, 1958a; *Evans*, 1969; *Zhou and Sulzer*, 1997; *Liu et al.*, 2007b). A radio signal is reflected at an altitude at which $f = f_N$ where f_N is the plasma frequency given by:

$$f_N = \frac{1}{2\pi}(\frac{Ne^2}{m_e\epsilon_0})^{\frac{1}{2}}, \tag{1.8}$$

where N is the electron number density in m^{-3}, e is charge on an electron, m_e is mass of an electron and ϵ_0 is permittivity of free space. f_N can be further reduced to

$$f_N = 8.97 \times 10^{-6} N^{0.5}. \tag{1.9}$$

Figure 1.5 shows a typical spectrum of the transmitted signal (i.e., the vertical black line) and the reflections (i.e., the gray under the red curve) of an incoherent scatter radar. The ionospheric parameters can be inferred by examining the received spectrum. First, the temperatures of the ionosphere can be determined by the width of the spectrum of the reflected signal. Because the transmitted signal is reflected by a very large number of particles which are moving in different directions with different speeds, the reflected signal composes a range of frequencies. Those reflected by particles moving away from the radar have reduced frequencies, while those reflected

by particles moving towards have increased frequencies. The electrons and ions usually have different temperatures. The width of the spectrum is a measurement of the ion temperature. The received spectrum usually has two shoulders, which are corresponding to electron temperature being raised higher than the ion temperature (*Evans*, 1969). Further, the number density of the electrons and ions can be measured by the strength of the received signal. The more charged particles are, the more power the received echo has. Moreover, the bulk velocity of the electrons and ions in the ionosphere can be measured by the shift of the center frequency between the transmitted and received spectrum.

The Global Positioning System (GPS) is widely used technique to measure the total electron content (TEC) (1 TECU = 10^{16} m^{-2}) in the ionosphere (*Mannucci et al.*, 1993; *Abdullah et al.*, 2009). TEC is the total number of electrons integrated along the path from a GPS to a receiver, which determines scintillation and group delay of a radio wave through the ionosphere. GPS satellites transmit electromagnetic waves on two frequencies: L1 (1575.42 MHz) and L2 (1227.60 MHz). A GPS receiver can measure the difference in the time delays between the L1 and L2 signals. The TEC can be obtained by

$$TEC = \frac{1}{40.3}\frac{f_1 f_2}{f_1 - f_2}(P_2 - P_1), \tag{1.10}$$

where P_1 and P_2 are the pseudoranges measured in L1 and L2, respectively (*Blewitt*, 1990). The pseudorange is expressed by

$$P_i = \rho + c(dT - dt) + \triangle_i^{iono} + \triangle^{trop} + b_i^{P,s} + b_i^{P,s} + m_i^P + \epsilon_i^P, \tag{1.11}$$

where i=1, 2 corresponding to carrier frequencies L1 and L2, ρ is the geometrical distance between the GPS satellite and receiver, c is the vacuum light speed, dT and dt are the receiver and satellite clock offsets from GPS time, $\triangle_i^{iono}$ is the ionospheric

delay, $\triangle^{trop}$ is the tropospheric delay, b_i is the receiver and satellite instrumental delay, m_i is the multipath on P measurement and ϵ_i is the receiver noise.

The slant TEC (TECs) is a measure of the TEC of the ionosphere along the signal path from the satellite to the receiver, which is dependent on the path geometry through the ionosphere, as shown in Figure 1.6. The vertical TEC (TECv), also known as a local TEC, depends only on geographical location and time. Assuming that the ionosphere is a spherical shell at fixed altitude, the slant TEC can be converted to vertical TEC using an obliquity correction factor:

$$TECv = TECs \cos \chi, \tag{1.12}$$

where χ is the zenith angle at the subionospheric point between the satellite and the receiver at a height given by the center of mass of the ionospheric profile, usually between 350 km and 450 km.

1.2 Dynamics in the upper atmosphere

1.2.1 High-latitude electrodynamics

The major mechanism of energy transfer from the solar wind to the Earth's magnetosphere and circulation in the coupled magnetosphere-ionosphere system is magnetic reconnection (*Dungey*, 1961). As shown in Figure 1.7, there are two reconnection locations in the magnetosphere: dayside of the magnetosphere or the magnetopause, and the cross-tail current sheet region. Except for the reconnection regions, the plasma in the magnetosphere and the F-region ionosphere are frozen to the magnetic field, i.e., the plasma moves together with magnetic flux tubes. When there is a southward component in the Interplanetary Magnetic Field (IMF), reconnection occurs at the magnetopause (i.e., line 1 in Figure 1.7), opening the geomagnetic field lines. The geomagnetic field lines are then connected to the IMF embedded in the solar wind

and are driven tail-ward by the solar wind plasma, as indicated by lines 2-5. These magnetic flux tubes are removed from the dayside and brought to the nightside. The open field lines are closed again by reconnection in the cross-tail current sheet region (i.e., line 6), and then transported sun-ward and back to the dayside. At the same time, the sudden release of power by the reconnection produces intense plasma injection into the inner magnetosphere, building up strong field-aligned currents and producing energetic particle precipitation in the upper atmosphere (*Frey et al.*, 2004).

The highly-conducting magnetic field maps the solar wind electric field down to the polar cap ionosphere. As shown in Figure 1.9, the imposed electric field drives an anti-sunward **ExB** plasma drift across the polar cap, and a sunward return flow at auroral oval latitudes (*Reiff*, 1982; *Heelis et al.*, 1982). The anti-sunward flow over the polar cap maps to the magnetospheric tail lobes, and the sunward return flow at lower latitudes maps mainly to the plasma sheet and ring current regions (*Ohtani*, 2000). Throughout the F-region, the ions are considered roughly collisionless with the neutrals and move with the **ExB** drift. Below the F-region, the amplitude of the ion drifts depend on the ratio of the collision frequency with neutrals to the ion gyro frequency. The ion-neutral collisional frequency, which is proportional to the neutral density, decreases exponentially with altitude. At an altitude of about 125 km, the ratio is approximately one. Above, the collision rate becomes small and the ion drift is considered to be **ExB**. In the region below ∼125 km, the **ExB** drift becomes negligible and the ion drifts approximately in the direction of **E**. At the altitude when the ratio equal to one, the ion drift is at a 45° angle between the **E** and **ExB** directions. However, since the electron-neutral collisional frequency is far less than the electron gyro frequency throughout the whole region of the ionosphere, the electrons drift in the **ExB** direction in both the E and F regions.

Because the electrons and ions can move in different directions, currents are generated: Pedersen currents are in the **E** direction and Hall currents are in the **ExB**

direction due to the electron motion. The Pedersen currents are dissipative, which converts electromagnetic energy to thermal energy of neutrals and ions. The Pedersen currents are divergent, and are connected to field-aligned currents. As shown in Figure 1.9, Region 1 field-aligned currents flow upward in the dusk sector and downward in the dawn sector at higher latitudes, while Region 2 currents are of the opposite direction. Region 1 currents are in the shear region between the sunward and anti-sunward plasma flow associated with the open-closed field line boundary (*Cowley and Lockwood*, 1992). The Region 1 and Region 2 currents are closed with Pederson currents (approximately). The Hall currents are divergence free (approximately) and mostly self-closed within the ionosphere.

1.2.2 Dynamics in thermosphere

Numerous observations revealed that there are three dominant external energy sources in the thermosphere: solar radiation, magnetospheric forcing, and lower atmosphere forcing (*Huba et al.*). The sun is one of the primary drivers of neutral density and temperature variability in the thermosphere, and it affects the thermosphere mainly from two aspects: solar radiation at EUV and soft X-ray wavelengths and geomagnetic disturbances associated with the interaction between the solar wind and the Earth's magnetosphere. Investigating the dynamics of the thermosphere (i.e., the thermospheric neutral density and temperature) due to solar forcing, is very important for understanding many space environment applications. For example, when the thermospheric density is enhanced during a solar flare or a geomagnetic storm, satellites at low-altitude orbits can experience an increase in satellite drag.

1.2.2.1 Solar radiation heating

Solar radiative heating is the dominant energy source into the thermosphere, which determines the basic structures and dynamics of temperature, density and composi-

tion of the thermosphere. The solar radio emission at a wavelength of 10.7 centimeters, known as the F10.7 index, has been found to be well correlated with the number of sunspots. A number of UV and EUV emissions that impact the Earth's upper atmosphere, are correlated with the F10.7 index as well. Sunspots are regions on the sun with an intense concentration of magnetic field flux and low plasma density and temperature. The variation of the sunspot number follows the 11-year solar cycle. The F10.7 radio emission can be easily obtained by ground-based measurements regardless of weather conditions, and has been widely applied for various scientific purposes. The recorded F10.7 index varies from approximately 65 solar flux unit (sfu) (1 sfu = 10^{-22} W m^{-2} Hz^{-1}) at solar minimum to a maximum value of 300 sfu at solar maximum (*SWPC*, 2015) as shown in Figure 1.8.

At a solar maximum, solar flux at wavelengths shorter than 200 nm (i.e., XUV, EUV, etc.) can cause significant disturbances in the space environment, such as modulations in atmospheric densities, compositions and ionospheric conductivity. Despite numerous measurements at this wavelength range, data gaps in frequency and time exists during during particular periods. To solve these issues, the Flare Irradiance Spectral Model (FISM) was developed by *Chamberlin et al.* (2007). FISM is an empirical model that estimates the solar irradiance at wavelengths from 0.1 to 190 nm resolution with a time cadence of 1 minute. Many Ionosphere/Thermosphere models have used the solar irradiance estimates by FISM to investigate the responses of the IT system to solar flares (*Qian et al.*, 2010, 2012; *Pawlowski and Ridley*, 2008).

The vertical profile of the global averaged thermospheric temperature is mainly determined by a balance between solar heating and thermal conduction cooling. The simplified form of the neutral energy equation can be written as (*Hedin and Mayr*, 1987):

$$-\frac{\partial}{\partial z}(\lambda_n \frac{\partial Tn}{\partial z}) = Q_{euv}, \tag{1.13}$$

where T is the global mean neutral temperature, z is altitude, λ_n is the neutral conductivity and Q_{euv} is the solar EUV heating. The solar EUV flux is absorbed by the neutral atmosphere through two channels: photoionization and photodissociation. Photoionization is a physical process in which an ion-electron pair produced from the interaction of a photon with an atom and molecule. The EUV energy is partitioned between the chemical energy of ions and kinetic energy of photoelectrons. These fast photoelectrons are slowed down by collision with thermal electrons or neutrals (*Torr et al.*, 1980a). The neutrals are then heated by collisions with ions and electrons. Photodissociation is a chemical reaction in which a molecule is broken down by photons, and the photon energy exceeding the reaction threshold heats the neutrals. These are the two main energy flow from solar EUV to the neutrals (*Schunk and Nagy*, 2009). The thermal conduction here acts as a cooling term, transporting heat to the lower atmosphere. Since the heat conductivity is density independent in the thermosphere, the temperature gradient is greater at the lower thermosphere (from roughly 90 km to 200 km) where the denser atmosphere absorbs the majority of the solar energy. The variation of solar EUV heating due the solar cycle ($\sim$11 years) and due to solar rotation of about 27 days also modulate the thermosphere status. The neutral density at 350 km can change by an order of magnitude over a solar cycle, and the neutral temperature can vary from about 700 K to 1200 K from solar minimum to solar maximum (*Hedin*, 1984). It was found that the atmospheric response at the long time scales (i.e., the 11 year solar cycle) is typically two orders of magnitude greater than it is at short time scales (i.e., 27 days solar rotation) due to the differences in variability at different solar radiative wavelengths (*Jacchia and Slowey*, 1973).

The neutral temperature in the thermosphere was found to be linearly correlated with the 10.7 index (*Covington*, 1948; *Schmahl and Kundu*, 1995). The dynamics of the thermosphere can be affected by solar activity in several ways. The solar

radiation is absorbed by the thermosphere on the dayside, which causes horizontal variations in temperature and density, and thus pressure gradient between the sunlit and the dark atmosphere. This drives a global neutral wind flowing from the dayside to the nightside. A weak return flow also exists in the lower thermosphere to close the circulation in order to maintain the mass continuity of the air flow. This is known as the solar-driven circulation. The meridional flow, which is pole-ward on the dayside and equatorward on the nightside, can be modified by the seasonal circulation. During summer, the sun shines continuously on the summer hemisphere, which causes the temperature in the summer hemisphere to increase significantly. This causes winds to flow away from the summer hemisphere and into the winter hemisphere. Also, the ion density increases due to an enhancement in photoionization rate during higher solar activities.

1.2.2.2 Magnetospheric forcing

Another important form of energy source of the thermosphere is Joule heating at high latitudes. Joule heating is one of the most important energy deposition processes from the magnetosphere into the upper atmosphere during storms and substorms, and it is strongly dependent on the convection electric field. The direct evidence that the neutral atmosphere can be affected by geomagnetic storms was obtained by *Jacchia* (1959), who detected changes in the decay rate of satellite orbits. Since this discovery, the global response of the thermosphere to magnetic storms and substorms were widely studied based on the air drag derived from the orbit decay of many different satellites (*Jacchia*, 1978; *Römer*, 1967). In the 1970s and early 1980s, satellite-borne gas analyzers (i.e., neutral mass spectrometers) have provided extensive information on the variations of the thermospheric composition during storms and substorms (*Carignan and Pinkus*, 1968; *Nier et al.*, 1973; *Bruinsma et al.*, 2006). Great efforts have been made to the understanding the physics of the thermospheric per-

turbations during magnetic disturbances using first-principle thermospheric models (*Killeen et al.*, 1997; *Thayer et al.*, 1995; *Deng and Ridley*, 2007). It has been shown that many physical and chemical processes, such as the electrodynamics, momentum and energy coupling between the neutrals and ions, and composition change due to chemical reactions, can affect the neutral atmosphere in complex ways (*Forbes*, 2007). A wave-like structure, which is known as a traveling atmospheric disturbance (TAD), has been both suggested by numerical simulations (*Richmond and Matsushita*, 1975; *Richmond*, 1979) and later detected by satellite measurements (*Gross et al.*, 1984; *Forbes et al.*, 1995). TADs are a spectrum of gravity waves generated by a sudden increase of high-latitude heating during magnetic disturbances. The waves are characterized by horizontal wavelengths of ∼1000-3000 km and a wave speed of about 1000-2000 m/s (*Bruinsma et al.*, 2004). The energy of TADs is mainly dissipated through collision with ions. Also, as shown in Figure 1.10, high-latitude heating can also drive a geomagnetic circulation system at high and middle latitudes, which moves against the background solar-EUV driven circulation.

The global circulation system is accompanied by mass transport and species redistribution in the thermosphere. The two major neutral species are O and N_2 below 500 km. Enhancements in N_2 at higher altitudes occur where increased magnetospheric energy input causes upwelling of molecular-rich gas from the lower atmosphere beyond the auroral oval (*Hays et al.*, 1973; *Bums et al.*, 1995). The nitrogen-rich air can then be redistributed by horizontal neutral winds, resulting in composition changes at mid-latitudes. The O/N_2 ratio, as shown by the labeled arrows in Figure 1.10, is closely coupled with negative storm effect (i.e., a decrease in the F_2 peak density (N_mF_2)) (*Prölss*, 1980). A high correlation has been found between the magnetic disturbance induced changes in the O/N_2 ratio and depletion of the ionospheric density in simulations and satellite observations (*Bums et al.*, 1995; *Immel et al.*, 2001; *Crowley and Meier*, 2008). The changes in the O/N_2 ratio is dependent on season,

latitude, local time and level of magnetic activity (*Rishbeth et al.*, 1987; *Rishbeth and Müller-Wodarg*, 1999). In the summer hemisphere, upwelling of molecular-rich air transports N_2 upward, leading to an increase in the neutral mean mass density and a decrease in the O/N_2 ratio. The nitrogen-rich air at higher altitude is transported into the winter hemisphere where there is no significant upwelling of air. The secondary circulation cell due to polar heating is associated with an upwelling and equatorward transport of neutrals. The interaction between the solar EUV and auroral driven circulations leads to a local maximum of the O/N_2 ratio near the boundary of the two circulation cells. It has been suggested that annual and semiannual, as well as longitudinal variations in N_mF_2 with quiet-day conditions are associated with the global circulation system (*Rishbeth et al.*, 2000; *Zou et al.*, 2000).

1.2.2.3 Thermospheric waves

Extensive wave activity exist in the thermosphere. The stratified structure of the atmosphere in combination with gravity leads to a restoring force on disturbed air parcels, which is known as the buoyancy force. When the wave frequency is significantly higher than the buoyancy frequency, the gravity effect can be ignored and the wave can be treated as ordinary acoustic waves. When the wave frequency is much lower than the buoyancy frequency, the kinetic and elastic energy of the acoustic waves are negligible, and the wave is known as gravity wave. When the wave frequency is comparable to the buoyancy frequency, both the potential energy and the kinetic energy may have identical effects. This is known as acoustic-gravity waves (*Press and Harkrider*, 1962; *Yeh and Liu*, 1974b). The acoustic and gravity branches are shown in Figure 1.11. The high-frequency branch is the acoustic branch and the low-frequency branch is known as the gravity branch. Here, ω_a is the acoustic cutoff frequency, $\omega_a = c_0/2H$ and ω_b is the buoyancy frequency, $\omega_b = (\gamma - 1)g^2/c_o^2$. c_o is the local sound speed and γ is the ratio of specific heats. The region $\omega_b < \omega < \omega_a$ is the

cutoff zone where no free waves exist.

Gravity waves were first detected as irregularities and irregular motion in the D, E and lower F region of the ionosphere (*Hines*, 1960), and have been studied by a variety of techniques such as meteor-trail photography and radar observations since the 1950s. Very strong small scale fluctuations, characterized by vertical wavelengths of a few kilometers and a velocity amplitude of tens of meters per second, were observed to superimpose on the general shear of the neutral wind profile. The horizontal characteristic size of the waves was 20 times greater than their vertical wavelengths. It was recognized that gravity waves played an important role in transporting energy and momentum, causing disturbances and mixing and modifying the global mean circulation system (*Mayr et al.*, 1990; *Oliver et al.*, 1997; *Fritts and Alexander*, 2003). Observations have revealed that the presence of gravity waves can extend to very high altitudes (i.e., 500 km and above) and could have important influences on the neutral dynamics on the ionospheric disturbances. Gravity waves are known as a major contributor to the traveling ionospheric disturbances, equatorial plasma bubbles formation and growth through neutral-ion coupling (*Huang et al.*, 1993; *Kherani et al.*, 2009). Large ionospheric density and velocity perturbations have been associated with gravity waves. Also, the observation of traveling ionospheric disturbances by ionosondes, GPS receivers, all-sky imagers and in situ measurements, were used as evidence of large gravity wave motion. The spatial and temporal scales of gravity waves in the mesosphere and lower thermosphere have been extensively studied (*Klostermeyer*, 1972; *Vadas and Fritts*, 2002). The amplitudes and spatial scales tend to increase with altitudes to maintain the conservation of energy.

There are a few mechanisms that can cause gravity waves in the thermosphere. The primary source occurs in the lower atmosphere under geomagnetic quiet conditions. Gravity waves launched by deep tropical convection can propagate upward and penetrate into the mesosphere and lower thermosphere. The vertical wavelength

can increase from zero near the ground to over hundreds of kilometers in the thermospheric altitude range (*Oliver et al.*, 1997). The other major source of internal gravity waves is body forcing dissipation and localized momentum flux divergence spatially and temporally, such as large energy and momentum input in the auroral zones during geomagnetic storms and sudden enhancement in solar irradiance during solar flares (*Fritts and Lund*, 2011).

1.2.3 Ionosphere thermal dynamics

Structures and variations of the Earth's ionosphere have been widely studied through ground-based and in situ measurement since 1960s. The first direct measurement of the electron temperature in the ionosphere was made by *Reifman and Dow* (1949) based on the Langmuir probe observations on a V-2 rocket. Following this effort, a great number of studies have been conducted based on rocket, satellite and ground-based observations (*Evans and Loewenthal*, 1964; *Evans*, 1965; *Farley et al.*, 1967; *Sayers*, 1970; *Willmore*, 1970). Considerable theoretical work has also been done on the thermal behavior of ionospheric plasma (*Rishbeth and Setty*, 1961; *Harris and Priester*, 1962; *Da Rosa*, 1966a; *Schunk and Nagy*, 1978b; *Schunk et al.*, 1986b).

Otsuka et al. (1998) reported a comprehensive picture of the diurnal, seasonal and solar-cycle temperature variations of the ionosphere from 200-550 km based on the middle and upper atmosphere radar observations during a nearly full solar cycle. They found that the electron and ion temperatures had similar diurnal and altitude variations but different seasonal and solar activity dependence. The electron temperature was characterized by morning and evening peaks, mainly due to a sudden enhancement in the photoelectron heating in the low-density electron gas. A strong inverse correlation between the electron temperature and density was observed above the Millstone Hill Radar (*Bilitza and Hoegy*, 1990; *Zhang and Holt*, 2004). They

suggested that with increasing electron density, the cooling terms exceed the heating terms and the temperature decreased, while with decreasing electron density, the heating dominated and the temperature increased. The anti-correlation is pronounced in winter and equinox but becomes weak or even disappears in summer. *Lei et al.* (2007) found two electron density thresholds at about 5×10^{11} and 13×10^{11} m^{-3}. The correlation between the electron density and temperature is negative when the electron density is between the two thresholds, and becomes positive below the lower or above the upper threshold.

Figure 1.12 shows the altitude profiles of the electron and ion temperatures observed by the Millstone incoherent scatter radar (*Zhang et al.*, 2004). The electron temperature in the ionosphere responds rapidly to changing external heating (in seconds). During the daytime, photoelectron heating is the dominant heat source for electrons. The steep gradient in T_e above 400 km is caused by a large downward electron heat flux from the inner magnetosphere. At night, due to the absence of solar heating, T_e is reduced to a comparable magnitude to neutral temperatures below about 250 km. An increased slope above 250 km, however, is an indication of the existence of a downward heat flux at the topside ionosphere. T_e is higher during low solar activity than it is during high solar activity. The solar activity level also affects the shape of the altitudinal profile of T_e. There is a peak around 250 km in T_e at solar minimum, while T_e progressively increases with height at solar maximum.

The ion temperature can be approximated by a balance among collisional coupling to hot thermal electrons and cold neutrals at middle latitudes, since the temporal variation, advection, adiabatic expansion and thermal conduction are negligible (*Schunk and Nagy*, 2009). At lower altitudes, cooling by the neutrals is dominant, so T_i is close to the neutral temperature both during the day and night. At higher altitudes, T_i is more coupled with T_e and increases with height.

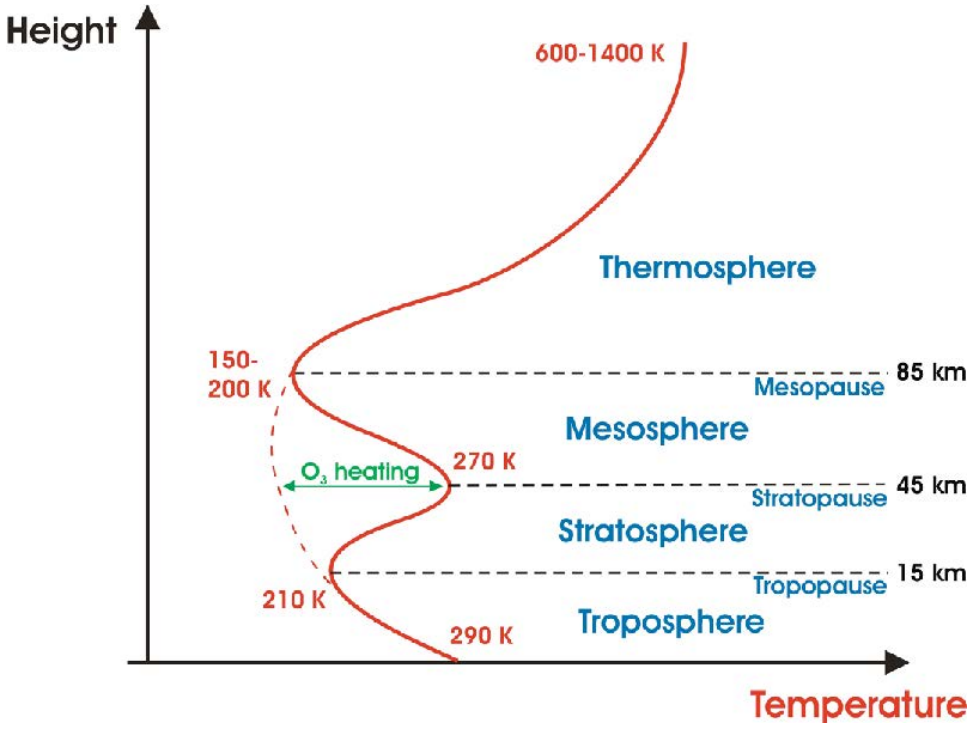

Figure 1.1: Vertical structure of the global mean temperature of the Earth's atmosphere. Courtesy of the Advanced School on Space Weather, ICTP, Itaty, 2006

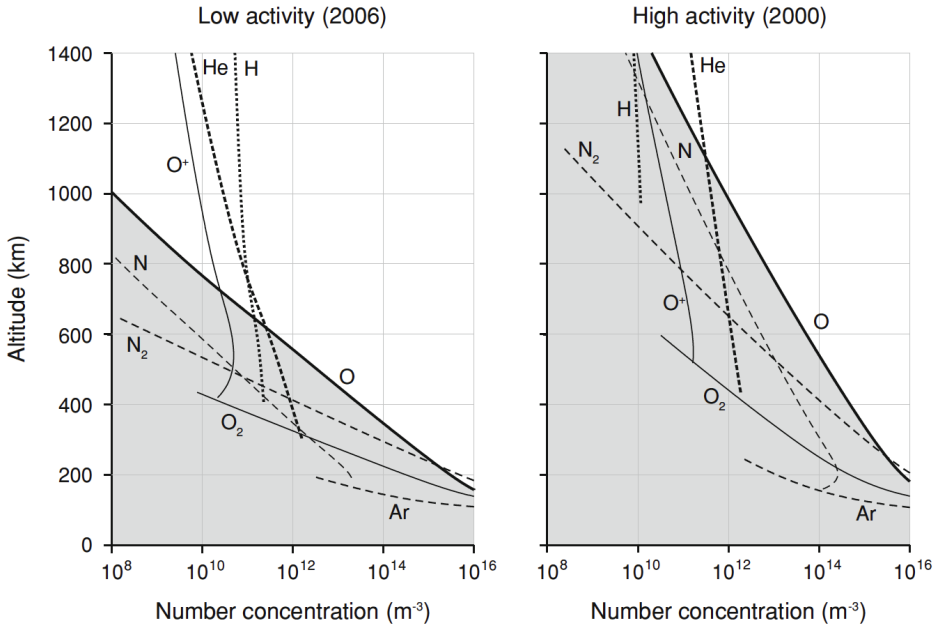

Figure 1.2: Altitude profiles of atmospheric number concentrations for low (left) and high solar activity levels, according to NRLMSISE-00 model, evaluated over Delft, at 18:00 on July 15, 2000 and 2006. Courtesy of UCAR Heliosphere VSP.

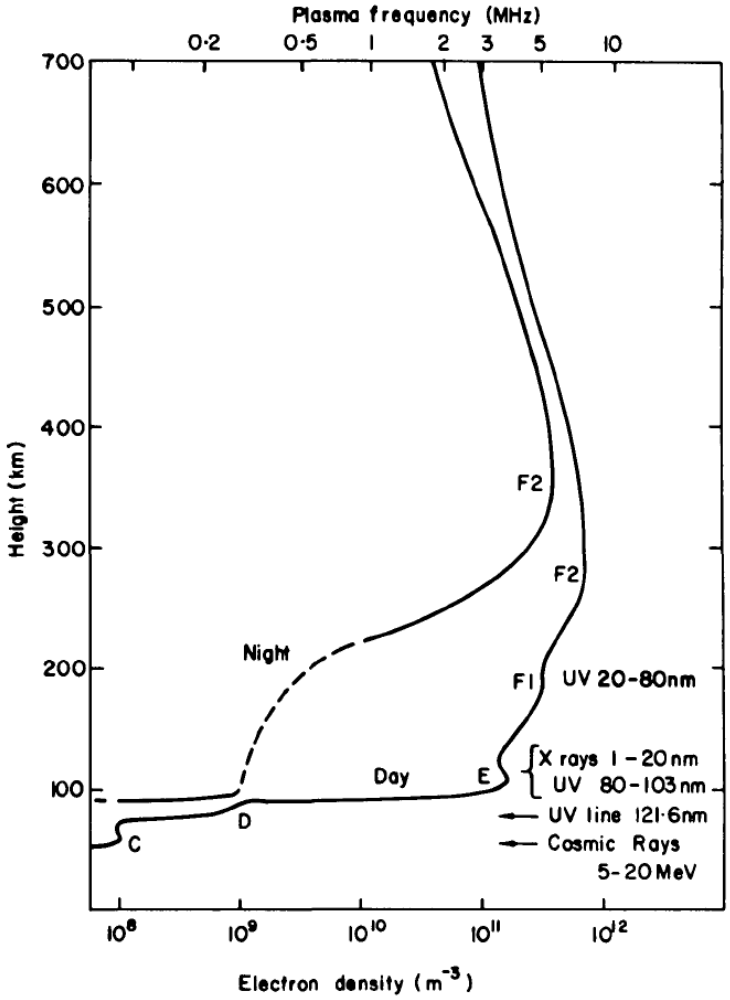

Figure 1.3: Typical mid-latitude electron density profiles for moderate activity, showing the radiations that produce the main ionospheric layers (*Rishbeth*, 1988).

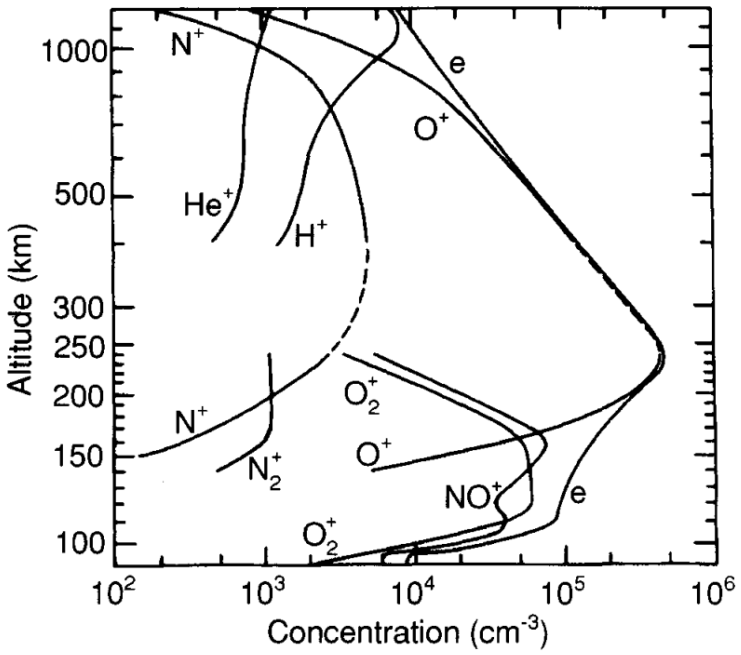

Figure 1.4: Typical composition of the dayside ionosphere at solar minimum (*Johnson*, 1966).

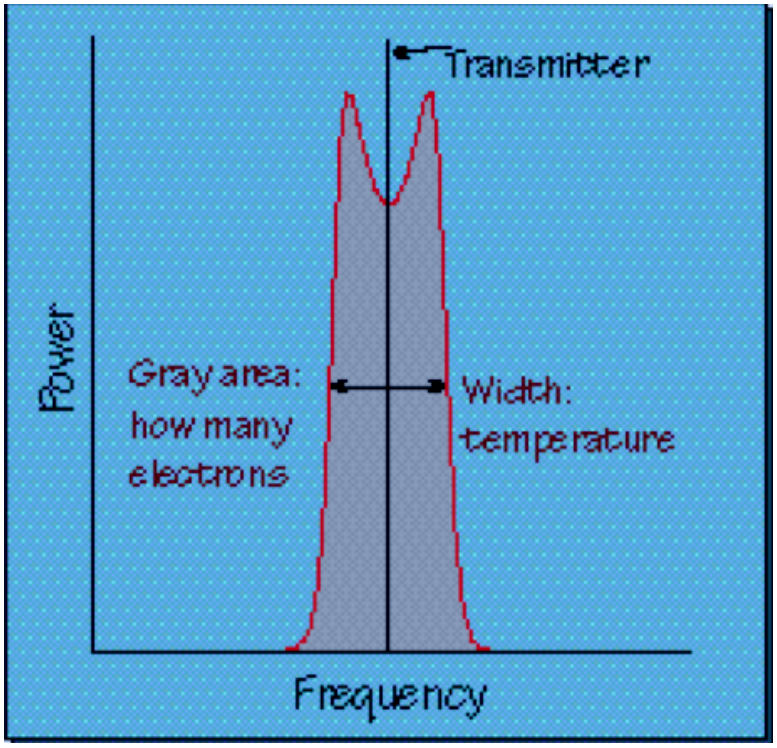

Figure 1.5: A typical spectrum of the transmitted signal and the reflections. Courtesy of the Arecibo Observatory.

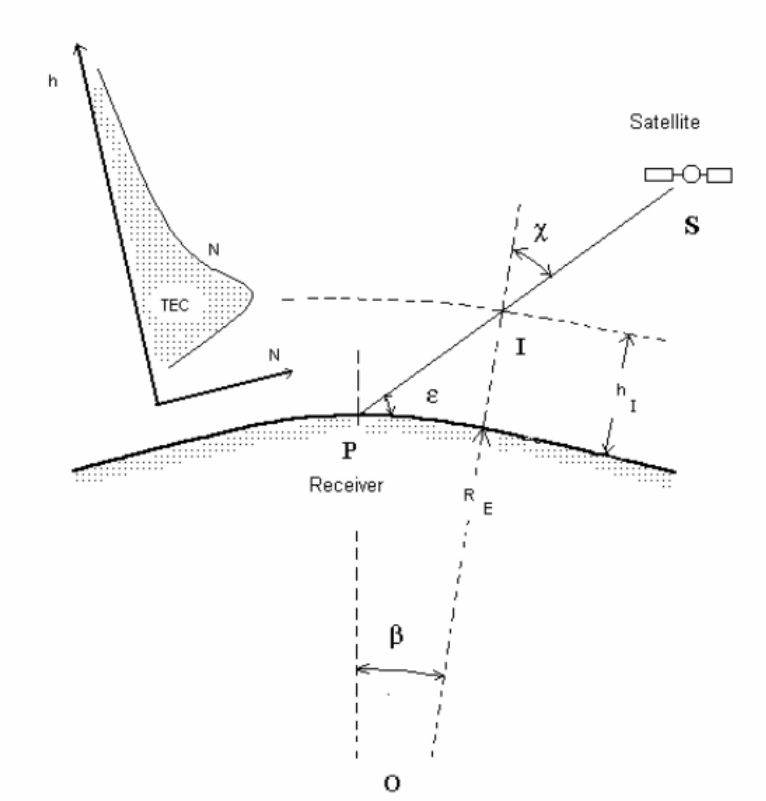

Figure 1.6: Schematic of vertical TEC mappling. Courtesy of I. J. Kantor et al.

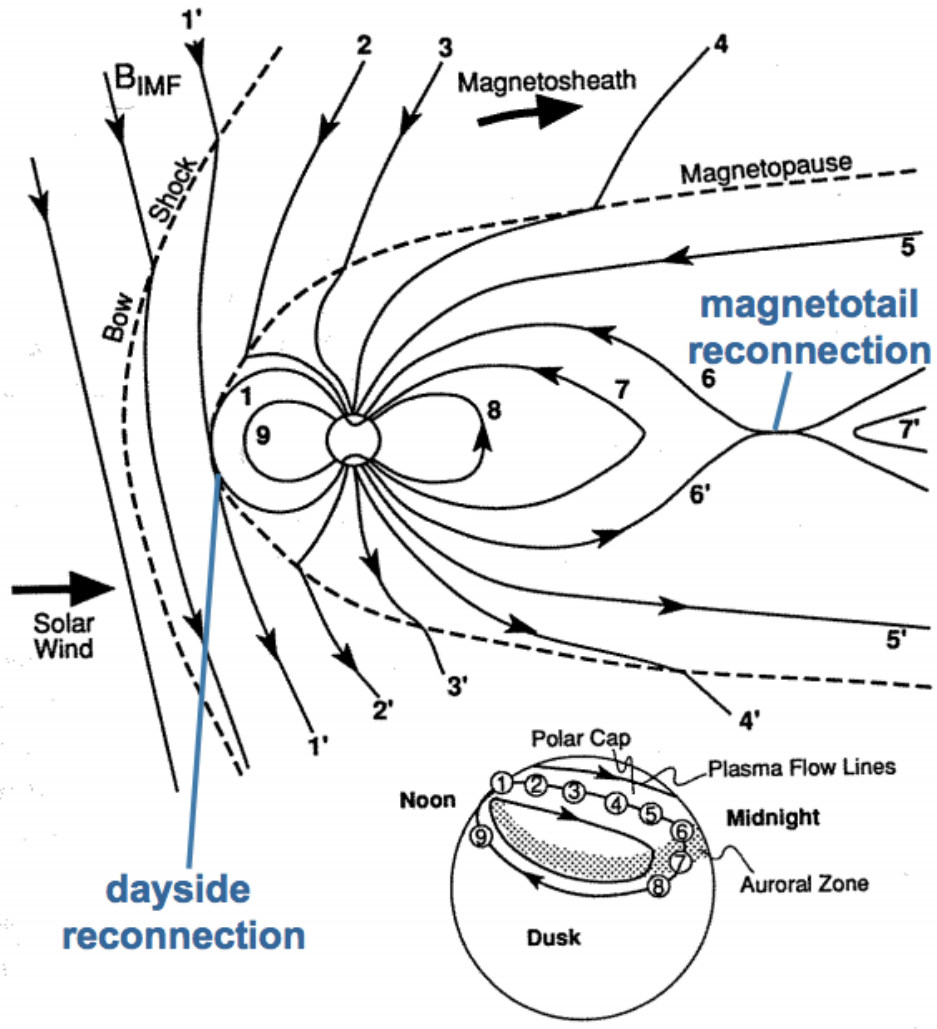

Figure 1.7: Schematic representation of a complete Dungey cycle. Courtesy from Introduction to Space Physics and Space Instrumentation at Gatech.

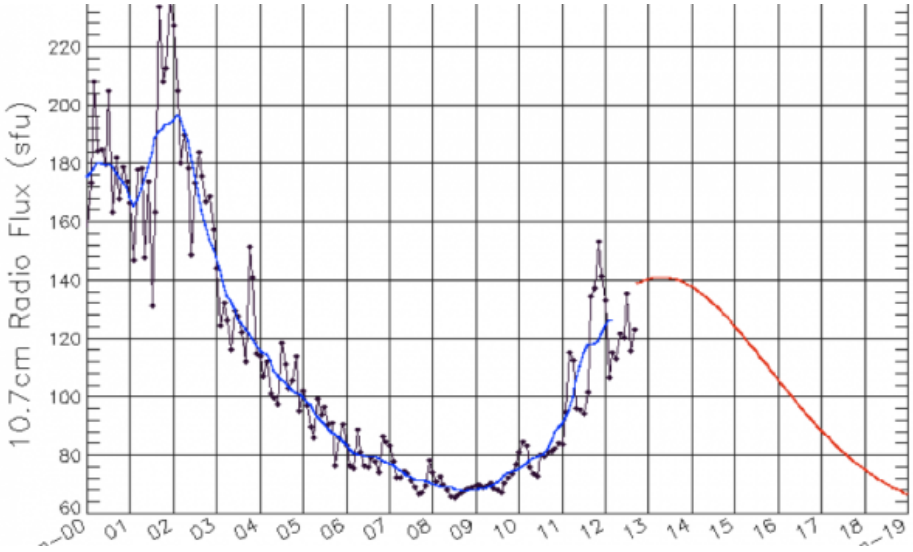

Figure 1.8: Yearly variation of F10.7 cm radio emissions. Courtesy of the Space Weather Prediction Center, NOAA.

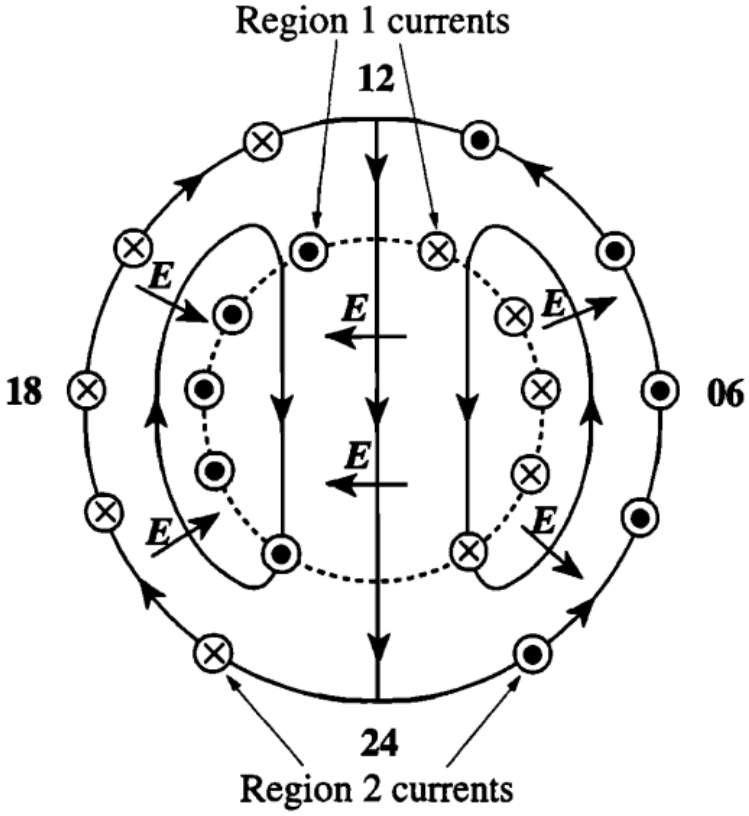

Figure 1.9: Schematic of Dungey-cycle flow mapped into the ionosphere, where the arrowed solid lines are the plasma streamlines, the short arrows give the direction of the electric field, and the dashed line is the open-closed field line boundary. The Hall current flows around the plasma streamlines opposite to the flow, while the Pedersen current flows in the direction of the electric field. The direction of the FAC flow associated with the horizontal divergence of the ionospheric currents is indicated by the circular symbols, where circled dots indicate upward currents out of the ionosphere, while circled crosses indicate downward currents into the ionosphere. (*Ohtani*, 2000)

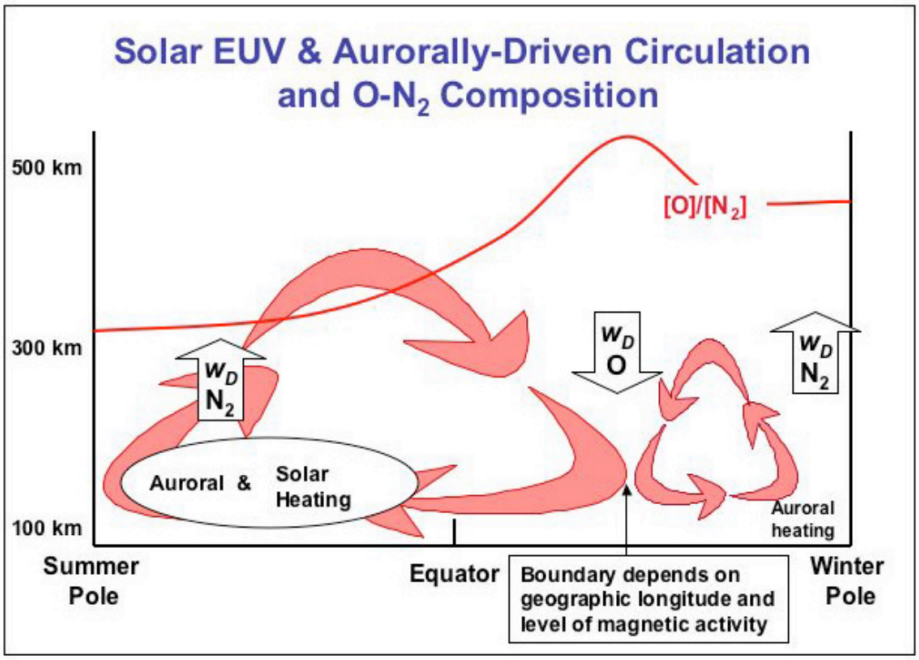

Figure 1.10: Schematic illustrating the zonal mean meridional circulation driven by differential solar heating in the thermosphere (blue arrows) and the transport of O and N_2 (labeled arrows), the latitudinal variations of R$[O/N_2]$, and a rough depiction of the boundary between mixed and diffusively-separated O-N_2 composition with estimates of the diffusive time constant (red). (*Forbes*, 2007)

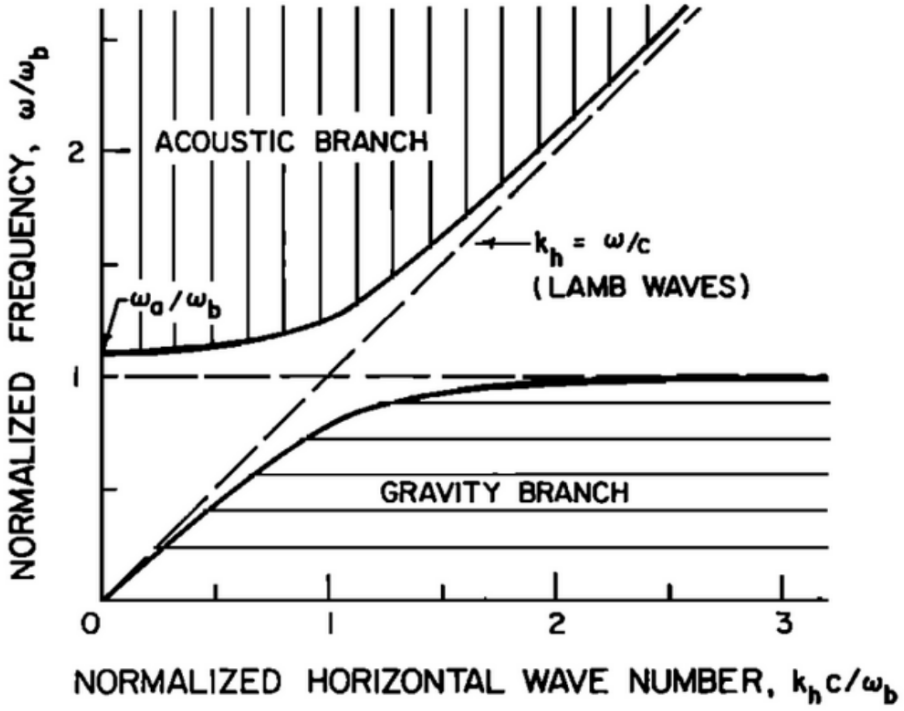

Figure 1.11: Regions of propagation of the gravity branch and the acoustic branch in an isothermal atmosphere. The bounding curves are given by k_z=0. The regions that the corresponding to finite real k_z values are shown by the hatched areas. (*Yeh and Liu*, 1974b)

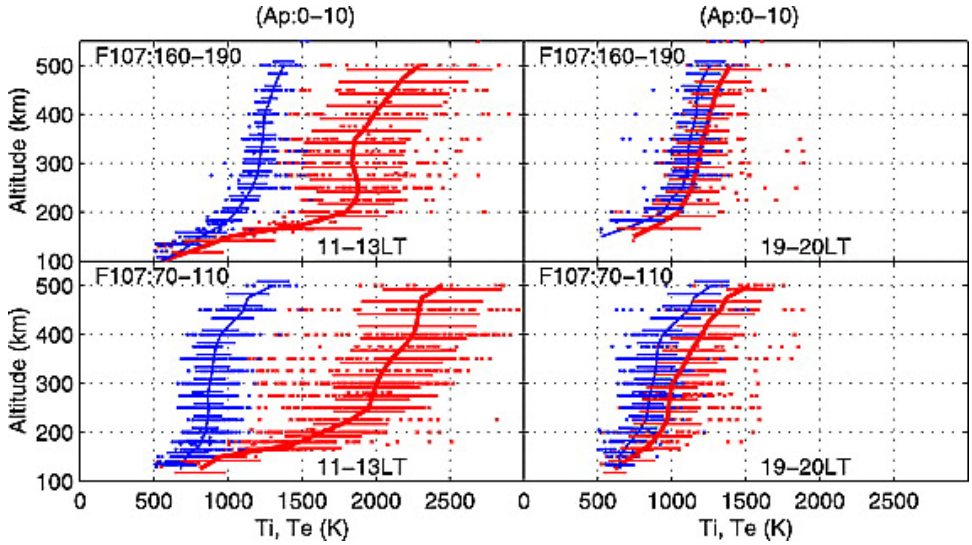

Figure 1.12: T_i and T_e altitude profiles at two local time spans for two solar cycle activity levels. The blue circles are for T_i and the red dots are for T_e. Solid lines for averages at various height bins. Error bars are for standard deviations from averages. (*Zhang et al.*, 2004)

CHAPTER II

Methodology

The main numerical tool used to investigate the effect of energy input of the upper atmosphere in this thesis is the Global Ionosphere Thermosphere Model (GITM). This chapter present a summary of the history of thermosphere-ionosphere models, and the basic physical equations and chemical processes in GITM.

2.1 History of thermosphere ionosphere modeling

Great effort has been made to simulate the Earth's upper atmosphere since the 1970s. Both empirical and physical models have been developed to investigate various phenomena and dynamics in the upper atmosphere for scientific purposes and to support many space weather applications.

One of the most important empirical models of the thermosphere is the Mass Spectrometer Incoherent Scatter (MSIS) model (*Hedin et al.*, 1977; *Hedin*, 1983, 1991). MSIS is a global model based on the OGO 6 Quadrapole Mass Spectrometer Model (*Hedin et al.*, 1974). It provides the thermospheric temperature, density and composition by using a spherical harmonic fit of comprehensive measurements from rocket flights, satellites and radar-based incoherent scatter measurements. The model includes local time, solar and geomagnetic activity, as well as UT/longitude dependences, and has been improved to extend into the mesosphere and lower atmosphere.

MSIS has been widely used because it provides a convenient way of calculating the average profile of the thermosphere for various user-specified geophysical conditions.

The International Reference Ionosphere (IRI) model is an important global empirical ionosphere model (*Bilitza*, 1986, 2001; *Bilitza and Reinisch*, 2008). IRI describes the monthly average of the global ionosphere in the range of 50-1500 km based on ground-based and in situ measurements over many decades. It has been shown that the mean value of the electron density in the F region could deviate by up to 25% compared to the measured value. Therefore, it is expected that some differences can exist between IRI and reality. IRI has been steadily improved and a model of the spread F occurrence probability and an auroral E-region electron density have been added to IRI-2007 (*Bilitza and Reinisch*, 2008). The IRI model has been used for a wide range of applications as a standard for the average ionospheric temperatures, densities and composition under specific solar and geomagnetic conditions.

Although empirical models provide great convenience in specifying the mean profiles of the upper atmosphere, it is difficult for them to examine the physics of and to predict the behavior of the thermosphere and ionosphere during various transient events. Physical models based on first principles have been developed since the 1980s. They solve the fluid equations self-consistently for the density, velocity and energy of the thermosphere and ionosphere. The thermosphere general circulation model (TGCM) is one of the first three-dimensional global numerical models of the thermosphere (*Dickinson et al.*, 1981; *Dickinson*, 1984). TGCM uses a fixed spatial resolution of 5° longitude by 5° latitude at 24 constant pressure layers. It uses global empirical models to specify a background ionosphere in order to sustain ion-neutral coupling processes, such as ion-neutral drag and chemical reactions. TGCM uses the neutral gas background from MSIS as the initial state and assumes that the thermosphere is in hydrostatic equilibrium. Two major energy sources are included in the model: solar EUV and UV radiation heating on the dayside and auroral heating

at high latitudes. A self-consistent global model of the coupled thermosphere and ionosphere (TIGCM) model was then developed by *Roble et al.* (1987). The model solves for the vertical profiles of O^+, NO^+, O_2^+, N_2^+ and N^+ assuming a photochemical equilibrium except for O^+ which has ambipolar diffusion included. *Roble and Ridley* (1987) also added a model of high latitude auroral processes in TGCM, which includes an empirical magnetospheric convection pattern (*Heelis et al.*, 1982) and an analytic prescription of particle precipitation in the auroral oval.

The thermosphere-ionosphere-electrodynamic general circulation model (TIEGCM) is a further improvement of TIGCM, which includes self-consistent electrodynamic interactions between the thermosphere and the ionosphere (*Richmond et al.*, 1992). TIEGCM calculates the dynamo electric field caused by the neutral wind, and then uses the resultant electric field and currents to calculate the dynamics of the neutrals and ions. The model is based on a realistic geomagnetic field, the International Geomagnetic Reference Field (IGRF) (*VanZandt et al.*, 1972). The geomagnetic field lines are assumed equipotential and currents flow along the field lines. The dynamo field is solved below about 65° magnetic field, and takes the potential from the *Heelis et al.* (1982) model along its polar cap boundary.

In order to investigate mesoscale processes in the global coupled ionosphere thermosphere system, a three-dimensional time-dependent nested grid model, Thermosphere-Ionosphere Nested Grid (TING) model was developed (*Wang et al.*, 1999). TING is an extension of the TIGCM. Specifically, TING includes two components: a regular "coarse" TIGCM and an adjustable nested grid code to simulate local structures of the thermosphere and ionosphere. The TIGCM and the nested code are coupled via a one way interaction scheme, which means that the output of TIGCM was taken as the initial and temporal boundary conditions for the nested grid, while the results of the nested grid do not affect the TIGCM outputs.

Another first-principle based numerical model is the Coupled Thermosphere Iono-

sphere Mode (CTIM) (*Fuller-Rowell and Rees*, 1980, 1983; *Fuller-Rowell et al.*, 1988), which solves for the fluid equations for the thermosphere and ionosphere self-consistently. *Millward et al.* (1996) further extended the model by adding a plasmasphere and low-latitude ionosphere to produce The Coupled Thermosphere Ionosphere Plasmasphere (CTIP) model. Later, the electrodynamics caused by neutral wind was also solved self-consistently for the neutral and plasma dynamics(CTIPe). The coupled middle atmosphere and thermosphere (CMAT) general circulation model is an extension of the CTIP model (*Harris et al.*, 2002). It covers an altitudinal range from about 30 km to about 300-600 km, and includes a full mesosphere-thermosphere neutral chemical scheme and gravity wave drag scheme.

The Global Assimilation of Ionospheric Measurement (GAIM) is a physics-based data assimilation model of the ionosphere and plasmasphere (*Schunk et al.*, 2004). It uses a Kalman filter to assimilate a diverse set of (near) real-time measurements from satellites, ionosode, occultation and ground-based GPSs. GAIM provides a three-dimension electron density profile from 90 km to a geosynchronous altitude of about 35,000 km. Also, drivers of the ionosphere, including neutral winds and densities, magnetosphere and dynamo electric fields and electron precipitation patterns, can be outputted by GAIM.

2.2 The Global Ionosphere Thermosphere Model

GITM is a three-dimensional model that simulates the Earth's coupled thermosphere and ionosphere from 100 km to about 600 km in spherical coordinates. The longitudinal and latitudinal resolution can be set flexibly. The spacing in altitude, as well as the lower and upper boundary of model the are set up based on the scale heights of the initial thermosphere. The entire domain in GITM is decomposed into blocks in longitude and latitude, which allows for fully parallel computation for each block. The information from neighboring blocks are exchanged using the Message

Passing Interface (MPI) standard.

GITM uses an altitude-based grid, which allows for non-hydrostatic solutions to exist in the upper atmosphere. Also, because it uses a small time step on the order of a few seconds, GITM is capable of resolving thermospheric waves in a variety of scales. This allows the model to capture non-hydrostatic wave activities that may occur when localized transient heating is present, i.e., solar flares or particle precipitation at high latitudes. The model solves for eight neutral species: O, O_2, N(^{2}D), N(^2P), N(^4S), N_2, NO and He. The continuity, momentum and energy equations are solved self-consistently for the neutrals and ions. Specifically, GITM solves for the bulk values for the neutral, ion and electron temperature, the horizontal neutral velocity, and the ion and electron velocities. The vertical momentum equations are solved individually for each neutral species, which allows a transition from a homosphere to a heterosphere in the thermosphere. In the homosphere, the Eddy motion is the primary diffusion process, where each species adopts the same scale height and the atmosphere is well mixed. In the lower thermosphere, the molecular diffusion starts to dominate where each species has its own scale height based on its mass. The transition height is called the homopause, and is where the Eddy diffusion coefficient and the molecular diffusion coefficient are equal. Heavier species have smaller scale heights, thus decrease faster with altitude. The ion momentum equation is solved in steady state and includes the following forces: gravity, gradient in pressure, neutral wind drag and external electric fields.

GITM allows flexible initialization and use of external drivers. The Earth's magnetic field can be represented either by an ideal dipole field or a more realistic field, i.e., IGRF. Also, different models of high-latitude potential, auroral particle precipitation and solar irradiance can be specified in GITM. The use of some important momentum and energy source terms can be turned on and off easily, which allows investigating the effect of individual source terms by conducting controlled experi-

ments. Moreover, the initial state can be set either by using the temperature and density from MSIS and IRI, or using the atmospheric status saved from a previous simulation.

2.2.1 Basic equations

GITM solves for the three-dimensional fluid equations in spherical coordinates. Because GITM does not assume a hydrostatic equilibrium, the neutrals are advected in both horizontal and vertical directions. Because the gradients in the vertical direction are much larger than they are in the horizontal directions, GITM solves for the vertical and horizontal advection separately. All the neutral species are advected with a bulk velocity horizontally, but with separate velocities vertically. Source terms are then added after the advection is completed.

First, the horizontal advection is presented, and all the source terms and terms containing vertical derivatives are neglected temporarily. In the horizontal direction, the bulk mass density is the sum of all the species densities, i.e.,

$$\rho = \sum_s M_s N_s, \tag{2.1}$$

where ρ is the total mass density, s is the index of a particular species, M_s and N_s are the molecular mass and number density of species s, respectively. The continuity equation is:

$$\frac{\partial N_s}{\partial t} + N_s \bigtriangledown \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \bigtriangledown N_s = 0, \tag{2.2}$$

where $\boldsymbol{u}$ is the bulk neutral velocity in the horizontal direction and t is time. Here, a normalized neutral temperature $\mathcal{T}$ is introduced:

$$\mathcal{T} = p/\rho, \tag{2.3}$$

where p is the total neutral pressure. Substituting $\mathcal{T}$ into the momentum equation results in:

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \bigtriangledown \boldsymbol{u} + \bigtriangledown \mathcal{T} + \frac{\mathcal{T}}{\rho} \bigtriangledown \rho = 0, \tag{2.4}$$

and the energy equation can be expressed as:

$$\frac{\partial \mathcal{T}}{\partial t} + \boldsymbol{u} \cdot \bigtriangledown \mathcal{T} + (\gamma - 1) \mathcal{T} \bigtriangledown \cdot \boldsymbol{u} = 0, \tag{2.5}$$

where γ is the specific heat ratio of a gas, i.e., the ratio of the specific heat at constant pressure to the specific heat at constant volume.

For the vertical advection, the logarithm of the total mass density ρ and the number density N_s are used as the primitive variables in the above equations:

$$\mathcal{R} = ln(\rho), \tag{2.6}$$

$$\mathcal{N}_s = ln(N_s). \tag{2.7}$$

This is because the density exponentially decreases with altitude and the logarithm of the vertical gradients varies linearly. Therefore, the logarithm of the density is much easier to solve numerically. Similarly, the normalized temperature $\mathcal{T}$ is used in the energy equation instead of p because p exponentially changes and is difficult to solve.

The vertical advection is treated individually for each species because the molecular diffusion becomes dominant and individual species start to decrease with their own scale height above about 120 km. The continuity equation for an individual species in the logarithmic form can be expressed as:

$$\frac{\partial \mathcal{N}_s}{\partial t} + \bigtriangledown \cdot \boldsymbol{u_s} + \boldsymbol{u_s} \cdot \bigtriangledown \mathcal{N}_s = 0. \tag{2.8}$$

The momentum equation for a species s can be written as:

$$\frac{\partial \boldsymbol{u_s}}{\partial t} + \boldsymbol{u_s} \cdot \bigtriangledown \boldsymbol{u_s} + \frac{\kappa}{M_s} \nabla T + \frac{\kappa}{M_s} T \nabla \mathcal{N}_s = 0, \tag{2.9}$$

where κ is the Boltzmann's constant, and T is neutral temperature, which is given by:

$$T = \frac{\bar{m}_n}{\kappa} \mathcal{T}. \tag{2.10}$$

Here, $\bar{m}_n$ is the number-density weighted averaged mass density, i.e.,

$$\bar{m}_n = \frac{\sum\limits_s M_s N_s}{\sum\limits_s N_s}. \tag{2.11}$$

In the energy equation, the bulk vertical velocity $\boldsymbol{u}$ is used, which is the average weighted by individual mass densities and can be expressed by:

$$\boldsymbol{u} = \frac{1}{\rho} \sum_s M_s N_s \boldsymbol{u_s}. \tag{2.12}$$

The equation sets are then projected into the rotating spherical coordinates with source terms added. The vertical continuity equation can be expressed as:

$$\frac{\partial \mathcal{N}_s}{\partial t} + \frac{\partial u_{r,s}}{\partial r} + \frac{2u_{r,s}}{r} + u_{r,s} \frac{\partial \mathcal{N}_s}{\partial r} = \frac{1}{N_s} \mathcal{L}_s, \tag{2.13}$$

where r is the radial distance from the center of the Earth, the subscript r denotes the radial component, $u_{r,s}$ represents the radial velocity of species s. $\mathcal{L}_s$ is the source term for neutral density for species s, which is given by:

$$\mathcal{L}_s = \frac{\partial}{\partial r}[N_s K_e(\frac{\partial N_s}{\partial r} - \frac{\partial N}{\partial r})] + C_s, \tag{2.14}$$

where K_e is the eddy diffusion coefficient, N is the total number density and C_s represents the sources or losses due to chemical reactions.

The momentum equation in the vertical direction in the spherical coordinates can be written as:

$$\frac{\partial u_{r,s}}{\partial t} + u_{r,s}\frac{\partial u_{r,s}}{\partial \theta} + \frac{u_\phi}{r\cos\theta} + \frac{\kappa}{M_s}\frac{\partial T}{\partial t} + T\frac{\kappa}{M_s}\frac{\mathcal{N}_s}{\partial r} = g + \mathcal{F}_s + \frac{u_\theta^2 + u_\phi^2}{r} + \cos^2\theta\Omega^2 r + 2\cos\theta\Omega u_\phi, \tag{2.15}$$

where θ is the north latitude, ϕ is the east longitude, u_θ is the northward neutral velocity, u_ϕ is the eastward neutral velocity, Ω is the angular velocity of the Earth, g is the gravitational acceleration, and $\mathcal{F}_s$ represents the forces due to ion-neutral and neutral-neutral friction. The last two terms on the right side are the centrifugal and Coriolis forces. The term $\frac{u_\theta^2+u_\phi^2}{r}$ results from the spherical geometry of the Earth, which means that horizontal winds tend to lift off from the curved surface of the Earth. $\mathcal{F}_s$ is given by:

$$\mathcal{F}_s = \frac{\rho_i}{\rho_s}\mu_{in}(v_r - u_{r,s}) + \frac{\kappa T}{M_s}\sum_{q\neq s}\frac{N_q}{N D_{qs}}(u_{r,q} - u_{r,s}), \tag{2.16}$$

where v_r is the vertical ion velocity, ρ_i is the total ion mass density, μ_{in} is the ion-neutral collision frequency, $u_{r,q}$ represents the vertical velocity of neutral constituents other than s, and D_{qs} is the diffusion coefficient between constituents s and q.

The vertical energy equation can be expressed as:

$$\frac{\partial \mathcal{T}}{\partial t} + u_r\frac{\partial \mathcal{T}}{\partial r} + (\gamma - 1)\mathcal{T}(\frac{2u_r}{r} + \frac{\partial u_r}{\partial r}) = \frac{\kappa}{c_v \rho \bar{m}_n}\mathcal{E}, \tag{2.17}$$

where c_v is the averaged specific heat at constant volume weighted by the number

density of each major species and $\mathcal{E}$ represents the energy source term. $\mathcal{E}$ contains solar EUV heating, NO and O cooling, thermal conduction and Joule heating, and can be written as:

$$\mathcal{E} = Q_{EUV} + Q_{NO} + Q_O + \frac{\partial}{\partial r}((\kappa_c + \kappa_{eddy})\frac{\partial T}{\partial r}) + N_e \frac{\bar{m}_i \bar{m}_n}{\bar{m}_i + \bar{m}_i} \mu_{in} (\boldsymbol{v} - \boldsymbol{u})^2, \quad (2.18)$$

where $\boldsymbol{v}$ is the ion velocity, μ_{in} is the ion-neutral collisional frequency, $\bar{m}_i$ and $\bar{m}_n$ are the average mass of ions and neutrals, respectively, and κ_c and κ_{eddy} are the thermal conductivity, and heat conductivity due to eddy diffusion, respectively. The thermal conductivity is given by:

$$\kappa_c = \sum_{s=O,O_2,N_2} \frac{N_s}{N_n} C_s T_n^{P_s}, \quad (2.19)$$

where s denotes individual species, N_s is the number density of species s, T_n is the neutral temperature in Kelvin, and N_n is the total number density of neutrals. C_s and P_s are constant coefficients for species s, and can be specified by the input file of GITM. The last term on the right side of Equation 2.18 is Joule heating which is due to the velocity difference between the ions and neutrals. Joule heating is one of the primary channels that transport energy from the Earth's magnetosphere to the upper atmosphere, leading to significant thermal and dynamic disturbances at high latitudes and even over the globe through wave propagation. An improvement of the Joule heating term will be provided in Chapter 6.

In the vertical direction, the continuity equation is given by:

$$\frac{\partial N_s}{\partial t} + N_s(\frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{1}{r\cos\theta}\frac{\partial u_\phi}{\partial \phi} - \frac{u_\theta \tan\theta}{r}) + \frac{u_\theta}{r}\frac{\partial N_s}{\partial \theta} + \frac{u_\theta}{r\cos\theta}\frac{\partial N_s}{\partial \phi} = 0. \quad (2.20)$$

The horizontal momentum equation is broken into two orthogonal directions: eastward (i.e., zonal) and northward (i.e., meridional). The eastward component of the momentum equation can be written as:

$$\begin{aligned}\frac{\partial u_\theta}{\partial t} + u_r\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{u_\theta}{r\cos\theta}\frac{\partial u_\theta}{\partial \phi} + \frac{1}{r\cos\theta}\frac{\partial \mathcal{T}}{\partial \theta} + \frac{\mathcal{T}}{r\rho\cos\theta}\frac{\partial \rho}{\partial \phi} \\ = \frac{\mathcal{F}_\phi}{\rho} + \frac{u_\theta u_\phi \tan\theta}{r} - \frac{u_r u_\theta}{r} + 2\Omega u_\theta \sin\theta - 2\Omega u_r \cos\theta,\end{aligned} \tag{2.21}$$

where $\mathcal{F}_\phi$ is the force due to ion-neutral friction and neutral viscosity in the longitudinal direction, and is given by:

$$\mathcal{F}_\phi = \rho_i \mu_{in}(v_\theta - u_\theta) + \frac{\partial}{\partial r}\eta\frac{\partial u_\theta}{\partial r}, \tag{2.22}$$

where η is the viscosity coefficient. The last two terms on the right side of Equation 2.21 are caused by Coriolis force.

The northward momentum equation can be written as:

$$\begin{aligned}\frac{\partial u_\theta}{\partial t} + u_r\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{u_\phi}{r\cos\theta}\frac{\partial u_\theta}{\partial \phi} + \frac{1}{r}\frac{\partial \mathcal{T}}{\partial \theta} + \frac{\mathcal{T}}{r\rho}\frac{\partial \rho}{\partial \theta} \\ = \frac{\mathcal{F}_\theta}{\rho} - \frac{u_\phi^2 \tan\theta}{r} - \frac{u_\theta u_\phi}{r} - \Omega^2 r\cos\theta\sin\theta - 2\Omega u_\phi \sin\theta,\end{aligned} \tag{2.23}$$

where $\mathcal{F}_\theta$ is the force in the northward direction, which contains the ion-neutral friction and neutral viscosity, and is expressed as:

$$\mathcal{F}_\theta = \rho_i \mu_{in}(v_\theta - u_\theta) + \frac{\partial}{\partial r}\eta\frac{\partial u_\theta}{\partial r}. \tag{2.24}$$

The last two terms on the right side of Equation 2.23 are the centrifugal and Coriolis forces, respectively.

The horizontal energy equation becomes:

$$\frac{\mathcal{T}}{\partial r} + \frac{u_\phi}{r\cos\theta}\frac{\partial \mathcal{T}}{\partial \theta} + (\gamma - 1)\mathcal{T}(\frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{1}{r\cos\theta}\frac{\partial u_\phi}{\partial \phi} - \frac{u_\theta \tan\theta}{r}) = 0. \tag{2.25}$$

Here, the energy source terms, which have been included in the vertical energy equation, are excluded.

The ion continuity equation is separated into the horizontal and vertical directions. The horizontal equation is:

$$\frac{\partial N_j}{\partial t} + \frac{v_\theta}{r}\frac{\partial N_j}{\partial \theta} + \frac{v_\phi}{r\cos\theta}\frac{\partial N_j}{\partial \phi} = \mathcal{L}_j, \tag{2.26}$$

where N_j is the number density of ion species j and $\mathcal{L}_j$ is the net source of the jth ion due to ionization and chemical reactions. Similar to neutrals, the logarithm of the density was solved for the ions in the vertical direction. The vertical continuity equation is thus given by:

$$\frac{\partial \mathcal{N}_j}{\partial t} + v_r\frac{\partial \mathcal{N}_j}{\partial r} + \mathcal{N}_j\frac{\partial v_r}{\partial r} = 0, \tag{2.27}$$

where $\mathcal{N}_j = ln(N_j)$. Note that the advection of the raw number density has also been implemented in GITM recently. In GITM, only O^+ is advected because O^+ is the dominant ion in the low-collision region (i.e., F-region) and other ion species are primarily controlled by local processes, such as ionization and chemical reactions.

The ion momentum equation can be expressed as:

$$\rho_i\frac{d\boldsymbol{v}}{dt} = -\bigtriangledown(p_i + p_e) + \rho_i\boldsymbol{g} + eN_e(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}) - \rho_i\mu_{in}(\boldsymbol{v} - \boldsymbol{u}), \tag{2.28}$$

where $\boldsymbol{v}$ is the ion velocity, p_i and p_e are the ion and electron pressure, e is the electron charge, N_e is the electron number density, $\boldsymbol{E}$ is the external electric field, and $\boldsymbol{B}$ is the

Earth's magnetic field. The temporal variation is assumed to be negligible compared to terms on the right side of the equation and the ion motion is assumed to be in steady state. Therefore, the ion momentum equation can be simplified to:

$$\boldsymbol{v} = \frac{\boldsymbol{A}}{\rho_i \mu_{in}} + \frac{eN_i}{\rho_i \mu_{in}}(\boldsymbol{v} \times \boldsymbol{B}), \tag{2.29}$$

where N_i is the total ion number density and $\boldsymbol{A}$ is given by:

$$\boldsymbol{A} = - \bigtriangledown (p_i + p_e) + \rho_i \boldsymbol{g} + eN_e \boldsymbol{E} + \rho_i \mu_{in} \boldsymbol{u}. \tag{2.30}$$

The ion velocity can be further solved to:

$$\boldsymbol{v} = \frac{\boldsymbol{A} \cdot \boldsymbol{b}}{\rho_i \mu_{in}} + \frac{\rho_i \mu_{in} \boldsymbol{A}_\perp + eN_i \boldsymbol{A}_\perp \times \boldsymbol{B}}{\rho_i^2 \mu_{in}^2 + e^2 N_i^2 B^2}, \tag{2.31}$$

where $\boldsymbol{b} = \boldsymbol{B}/B$ is the unit vector of the magnetic field, B is the magnitude of the magnetic field, and $\boldsymbol{A}_\perp$ is the component of $\boldsymbol{A}$ perpendicular to the magnetic field. The first term on the right-hand of Equation 2.31 represents forcing along the magnetic field, while the last term is the velocity component perpendicular to the magnetic field.

The electron number density is equal to sum of the number densities of ions since the ionosphere is charge-neutral. The motion of the electrons is mainly controlled by the $\boldsymbol{E} \times \boldsymbol{B}$ drift, i.e.,

$$\boldsymbol{v}_e = \frac{\boldsymbol{E} \times \boldsymbol{B}}{B^2}. \tag{2.32}$$

where $\boldsymbol{v}_e$ is the electron velocity.

The electron energy equation is given by:

$$\frac{\partial T_e}{\partial t} + \frac{2}{3} T_e \bigtriangledown \cdot \boldsymbol{v}_e - \boldsymbol{v}_e \cdot \bigtriangledown T_e + \frac{2}{3} \frac{1}{N_e \kappa} (- \bigtriangledown \cdot \boldsymbol{q}_e + Q_e + L_e), \tag{2.33}$$

where T_e is the electron temperature, N_e is the electron density, Q_e and L_e are the sum of the electron heating and cooling rates, respectively. The first to third terms on the right-hand of the equation represent the adiabatic heating, heat advection and the heating due to electron heat flow $\boldsymbol{q}_e$. The heat flow vector is given by:

$$\boldsymbol{q}_e = -\beta_e \boldsymbol{J}_\perp - \kappa_e \bigtriangledown T_e, \tag{2.34}$$

where β_e is the thermoelectric coefficient, $\boldsymbol{J}_\perp$ is the current parallel to the magnetic field, and κ_e is the electron thermal conductivity. The implementation of the electron and ion temperatures in GITM will be discussed in detail in Chapter 5.

2.2.2 Solar irradiance model

The incident extreme ultraviolet radiation is the dominant energy source into the upper atmosphere. GITM includes several models of solar irradiance: the *Hinteregger et al.* (1981) Solar EUV flux model (SERF1), the updated *Tobiska and Barth* (1990) (SERF2) model, the EUVAC model (*Richard et al.*, 1994; *Richards et al.*, 1994), the EUV spectrum provided by the Solar EUV Experiment (SEE) on the NASA Thermosphere, Ionosphere, and Mesosphere Energetic and Dynamics (TIMED) mission (*Woods et al.*, 1998), and the Flare Irradiance Spectral Model (FISM) (*Chamberlin et al.*, 2007, 2008; *Pawlowski and Ridley*, 2008).

The first two models, SERF1 and SERF2, provide EUV spectrum based on the daily averaged $F_{10.7}$ index and the averaged $F_{10.7}$ over the 81 days. The SERF2 model uses the EUV dataset measured by Atmosphere Explorer E (AE-E) satellite and 18 separate rocket flights for all levels of solar activity. Moreover, the H Lyman α line at 121.6 nm observed by the Solar Mesosphere Explorer satellite is also included. The model includes 39 wavelength bins or discrete lines from 1.9-105 nm.

The EUVAC model provides flux in 37 wavelength bins from 5-105 nm based on the measured F74113 solar flux. The flux is also scaled for solar activity by using

the $F_{10.7}$ index and its 81-day average. In GITM, the EUVAC model was modified to provide 59 wavelength bins from 0.1 nm to 175 nm.

The above models provide the EUV flux on a daily basis. However, in order to capture solar flux variations on a smaller time scale, more realistic EUV models are needed. The EUV measurement by SEE on aboard TIMED can also be used in GITM. SEE approximately provides one observation every 90 minutes and it contains spectral irradiance from 0.1 to 200 nm with about 0.5 nm spectral resolution.

During a solar flare, the soft X-rays solar flux can increase by two orders of magnitude in about 10 minutes (*Woods et al.*, 2004) and a flare typically lasts for a couple of hours. This indicates that SEE cannot provide enough measurements of solar irradiance during a solar flare. *Pawlowski and Ridley* (2008) used the interpolated SEE data by assuming an exponential decay during the recovery phase of a solar flare. Due to the sparsity of the SEE data, the model used the NOAA Geostationary Operational Environmental Satellites (GOES) data to determine the time of the flare onset. GOES provides X-ray spectrum from 0.5-4 and 1-8 Angstrom wavelength bins every one minute.

A more sophisticated model of interpolating the SEE data is to use the X-ray data measured by GOES (*Chamberlin et al.*, 2007) in FISM. Six different proxies are used to calculate the solar irradiance variations: the $F_{10.7}$ index representing the coronal continuum emission, the XUV irradiance integration from 0-4 nm from the SEE instrument, the 1 nm bin centered on 36.5 nm from SEE which has a strong center-to-limb brightening effect, the 1 nm bin centered at 30.5 nm representing the upper transition region emissions, the neutral hydrogen Lyman-α emission line at 121.6 nm, and the Mg II core-to-wing ratio representing the chromospheric emissions. FISM estimates the solar irradiance at wavelengths from 0.1 to 190 nm at a spectral resolution of 1 nm with a time cadence of 60s. Therefore, FISM is capable of modeling the variation of the solar flux during a solar flare, variations in a solar cycle and a

solar rotation.

The EUV heating absorbed by the thermosphere at a certain wavelength at a specific solar zenith angle and at a specific radial distance is given by the Chapman's incidence integral (*Chapman*, 1930; *Smith and Smith*, 1972):

$$Q_{EUV}(\lambda) = \epsilon I_\infty e^{-\tau(z,\lambda,\chi)} \sum_s \sigma_{abs,s}(\lambda) n_s, \tag{2.35}$$

where λ is the wavelength, z is the altitude, s denotes the neutral species, ϵ is the EUV heating efficiency of the thermosphere, I_∞ is the solar flux at the top of the Earth's atmosphere, τ is the optical depth, χ is the solar zenith angle, σ_{abs} is the absorption cross section, and n is the number density. The total solar heating of the thermosphere at each grid point is equal to the sum of the EUV heating over the entire wavelength range. The EUV heating efficiency represents the fraction of the absorbed EUV energy heating the thermosphere. *Stolarski et al.* (1975) showed that the neutral heating efficiency was between 0.3 and 0.35 below 400 km. *Torr et al.* (1980b) further showed that the heating efficiency was maximized at about 0.5 near the altitude when EUV irradiance deposits most of the energy, and decreased to about 0.1 near 400 km. Note that the neutral heating efficiency discussed above includes both the direct and indirect heating due to chemical reactions. Since the indirect heating was not self-consistently calculated in the model, a constant value of 0.05 is used in GITM. This value has been used to calculate the thermal structure of the thermosphere by *Roble et al.* (1987); *Stolarski et al.* (1975). The energy coupling between electrons, ions and neutrals has been recently added in GITM (*Zhu and Ridley*, 2015). It has been found that by adding the indirect heating of electrons to neutral through ions, the neutral heating efficiency can be further reduced to 0.01.

2.2.3 Chemistry

Chemical reactions dominate local sources and losses of ions in the ionosphere, as well as providing significant heating to the neutral atmosphere. In GITM, chemical reactions between ions and neutrals, ions and electrons, and neutrals and neutrals are included. The primary ion source on the dayside is the photoionization due to the solar EUV flux. The dominant photoionization reactions in the ionosphere are:

$$O + h\nu \rightarrow O^+(^4S) + e, \tag{2.36}$$

$$O + h\nu \rightarrow O^+(^2D) + e, \tag{2.37}$$

$$O + h\nu \rightarrow O^+(^2P) + e, \tag{2.38}$$

$$O_2 + h\nu \rightarrow O_2^+ + e, \tag{2.39}$$

$$N_2 + h\nu \rightarrow N_2^+ + e, \tag{2.40}$$

$$N + h\nu \rightarrow N^+ + e. \tag{2.41}$$

The major charge exchange reactions are:

$$O^+ + N_2 \rightarrow NO^+ + N, \tag{2.42}$$

$$O^+ + O_2 \rightarrow O_2^+ + O, \tag{2.43}$$

$$N_2^+ + O_2 \rightarrow O_2^+ + N_2, \tag{2.44}$$

$$N_2^+ + O \rightarrow O^+ + N_2, \tag{2.45}$$

$$N_2^+ + O \rightarrow NO^+ + N. \tag{2.46}$$

The primary dissociative recombination reactions are:

$$NO^+ + e \rightarrow N + O, \tag{2.47}$$

$$O_2^+ + e \rightarrow O + O, \tag{2.48}$$

$$N_2^+ + e \rightarrow N + N. \tag{2.49}$$

All the chemical reactions in *Rees* (1989) are included in GITM.

GITM uses the ionization rates due to auroral electron precipitation described by *Frahm et al.* (1997). The partitioning of the auroral ionization rates are given by *Rees* (1989):

$$P_{O^+} = 0.56 P_A \frac{[O]}{0.92[N_2] + [O_2] + 0.56[O]}, \tag{2.50}$$

$$P_{O_2^+} = P_A \frac{[O_2]}{0.92[N_2] + [O_2] + 0.56[O]}, \tag{2.51}$$

$$P_{N_2^+} = 0.92 P_A \frac{[N_2]}{0.92[N_2] + [O_2] + 0.56[O]}, \tag{2.52}$$

where P_A is the total ionization rate due to particle precipitation, P_{O^+}, $P_{O_2^+}$ and $P_{N_2^+}$ are the auroral ionization rate for O^+, O_2^+ and N_2^+, respectively, and $[O]$, $[O_2]$ and $[N_2]$ are the number density of O, O_2 and N_2, respectively.

GITM uses a sub-cycling technique to calculate the chemical reactions, which is more accurate than an implicit scheme since it does not require any assumption on chemical equilibrium. The sub-cycling is performed on a grid-cell basis. At altitudes below about 120 km (i.e., the E-region), the chemical time step could be reduced to below 0.1 second, while above about 150 km (i.e., the F-region), the chemical time step is similar to the advection time step, i.e., 2-4 seconds. The sub-chemical time step is determined by enforcing that the number density of any species can not be reduced by 25% at maximum. If the number density is decreased by less than 25% in an advection time step, only one iteration is conducted. If the number density is decreased by more than 25%, multiple sub-chemical time steps are taken until the percentage change is less than 25%. The number of chemical iterations is determined by the advection time step divided by the chemical time step. Typically, only one sub-chemical time step is needed above about 150 km, while approximately 2000 sub-cycling time steps are needed below around 110 km. By using the sub-cycling technique, GITM does not have to assume a local chemical equilibrium, and is capable of being used to investigate the chemical dynamics precisely.

CHAPTER III

The effect of background conditions on the ionospheric response to solar flares

The ionospheric response to two X5 solar flares that occurred in different seasons was investigated using the global ionosphere-thermosphere model (GITM). Two questions were investigated: (a) how do different solar flares with similar X-ray peak intensities disturb the ionosphere during the same background and driving conditions? and (b) how do the geomagnetic field and season affect the ionospheric response to solar flares? These questions were investigated by exchanging the two X5 flares for each other, so that there were two pairs of flares with (1) the same background conditions but different irradiances and (2) different background conditions but the same irradiance. The simulations showed that the different solar flares into the same background caused ionospheric disturbances of similar profiles but different magnitudes due to differences in the incident energies, while the same flare spectra caused perturbations of similar magnitudes but different TEC profiles in different backgrounds. On the dayside, the response is primarily controlled by the total integrated energy of the flare, independent of the background. For the northern and southern polar regions, the response is strongly controlled by the solar zenith angle and the incident energy, while the background plays a secondary role. On the nightside, the background conditions, including the magnetic field and season, play a primary role, with the neutral

winds and electrodynamics driving the ionospheric response.

3.1 Introduction

Interest in the influence of large solar flares on the Earth's ionosphere has gradually increased over the last few decades due to its deleterious effect on radio wave communication and navigation (*Garriott et al.*, 1967; *Davies*, 1990). Enhanced X-ray and Extreme-ultraviolet (EUV) irradiance during a solar flare causes increased ionization in the lower ionosphere (D and E regions) all the way up to the F region, depending on the flare spectrum. The increase of electron density in the F region is responsible for the increased total electron content (TEC) (*Mendillo et al.*, 1974). This phenomenon, known as a "sudden ionospheric disturbance" (SID) has been extensively studied for several decades (*Donnelly*, 1967; *Jones*, 1971; *Stonehocker*, 1970). Measurements from the ground-based world-wide Global Positioning System (GPS) network have recently been used to map TEC globally (*Coster and Komjathy*, 2008; *Rideout and Coster*, 2006). The sudden increase of TEC has been found to be linearly related to the cosine of the solar zenith angle (SZA) (*Zhang and Xiao*, 2003, 2005a). It was also found that both the flare-induced TEC variation rate, which was derived by differencing the vertical TEC measured by a station-satellite pair at continous epochs, was proportional to the increase in irradiance caused by the flare and inversely proportional to the Chapman function (*Wan et al.*, 2005). Based on the GPS measurements, an ionospheric solar flare activity indicator was given by the solar zenith angle and the TEC variation rate (*Xiong et al.*, 2014). The solar EUV flux, instead of the X-ray flux, was found to be primarily responsible for the increased TEC during and immediately after solar flares (*Tsurutani et al.*, 2009). The X17 flare on Oct 28, 2003 increased the TEC of the subsolar ionosphere by up to 30% in about 5 minutes (*Tsurutani et al.*, 2005). By using the effect of partial shadowing of the atmosphere by the Earth, the contribution from different ionospheric regions

to the TEC enhancements caused by the July 14, 2000 solar flare was estimated by *Leonovich et al.* (2002). They found about 20% of the TEC increase correspond to the ionospheric region lying below 100 km, about 5% of the increase came from the E region (100-140 km), about 30% came from the F1-region (140-200 km) and about 30% from the region above 300 km (*Leonovich et al.*, 2002). *Xiong et al.* (2011) found that the electron density in the E region was greater than that in the F region during the limb solar flare on September 7, 2005, which was mainly attributed to weak enhancements in the EUV flux and strong enhancements in the X-ray flux during this flare.

Despite the fact that GPS measurements have provided an effective way of detecting slant TEC disturbances due to solar flares, it is still difficult to observe the global ionospheric disturbance in detail, due to large gaps in the GPS station coverage around the globe. With the development of global models of the upper atmosphere using high-resolution solar spectra, the ionospheric response to dynamic solar forcing can be investigated more completely than before. Models also provide ways to probe the effects of individual driving forces, which is difficult to do with nature. The National Center for Atmospheric Research (NCAR) Thermosphere-Ionosphere-Mesosphere Electrodynamics General Circulation Model (TIME-GCM) has been used to investigate how the location of a solar flare on the solar disk affects the thermospheric and ionospheric response. *Qian et al.* (2010) showed that flare-driven changes in the F-region, TEC and neutral density in the upper thermosphere are 2-3 times stronger for a disk-center flare than a limb flare, due to the importance of the EUV enhancement. These model results are in agreement with the experimental results presented by *Leonovich et al.* (2010). It was also found that solar flares can impact the ionospheric electrodynamics by weakening the upward E×B drift in the magnetic equatorial region during solar flares. Large enhancement of E-region electron density increases both Hall and Pedersen conductivities, and increases the field-line

integrated Hall-to-Pedersen conductivity ratio at low latitudes, which leads to a decrease in the eastward electric field. The weakening of the upward E×B drift results in a decreased height and a reduced electron density of the F_2 peak (*Qian et al.*, 2012). Significant and long-lasting perturbations in TEC on the nightside have been shown in simulations by GITM (*Ridley et al.*, 2006) due to dynamical changes in the neutral atmosphere (*Pawlowski and Ridley*, 2009).

The study presented here explored how the background geomagnetic and season affects the ionospheric response to different solar flares with similar flare classes by using a global ionosphere thermosphere model. Specifically, two questions were examined: (a) What causes the ionosphere to react differently to different flares that have similar peak magnitudes? and (b) How different would the ionospheric response be to the same flare if the flare took place in a different season or with a different magnetic field configuration?

3.2 Model

To address these questions, GITM was used. In this case, the (*Weimer*, 2005) model was used to specify the high latitude electric field, a statistical model of electron precipitation based on hemispheric power was used to specify the aurora (*Fuller-Rowell and Evans*, 1987) and an empirical model of the thermosphere was used to specify the tidal structure of the neutral atmosphere just below 100-km altitude (*Hedin*, 1987). For all simulations described below, unless otherwise specified, IGRF was used to describe the magnetic topology. The solar irradiance spectrum calculated by FISM (*Chamberlin et al.*, 2007) was used to drive the model.

3.3 Methodology

This study explores the response of the ionosphere to two X5 flares. It was speculated that the response may depend on both the flare spectrum and the thermospheric and ionospheric background conditions when the flare occurred. In order to explore these dependences, multiple simulations of the coupled system were conducted. Each simulation was carried out using different conditions in order to determine the role of that particular condition on the ionospheric response to a solar flare.

The two flares that were explored were those on July 14, 2000 and April 6, 2001. These two flares had almost identical classes: X5.7 and X5.6 respectively. Solar flares are classified according to the peak brightness in the 0.1-0.8 nm X-ray wavelength range. This means that the two flare should have nearly identical peak intensities in X-ray wavelengths, but could have different brightnesses in EUV wavelengths.

Figure 3.1 shows the irradiances at different wavelengths for the flares in July (left) and April (right). The top row (a) shows the 0.1-0.8 nm X-ray irradiance for three days to provide a long-term context to the flares. Rows 2-4 show the irradiances in 0.1-0.8 nm (b), 58.43 nm (c) and 97.7 nm (d) surrounding the flares. The red lines show the 12-hour-medianed irradiances at corresponding wavelengths. The two EUV wavelengths, 58.43 nm and 97.7 nm, were plotted because the large cross section of atomic Oxygen at 58.43 nm and of molecular Oxygen and molecular Nitrogen near 97.7 nm (972.5 nm) (*Huffman*, 1969). The irradiances in 0.1-0.8 nm increased by a factor of approximately 40 and showed identical peaks for the two flares. The July flare lasted approximately twice as long above $10^4 W/m^2$ than the April flare. The irradiance at 58.43 nm increased by almost a factor of two for both flares. The April flare peaked slightly higher but with a shorter duration than the July flare. At 97.7 nm, the duration of the two flares were similar and the peak of the April flare was approximately one third higher than the July flare. The duration of the flare decreased with increasing wavelengths.

The July flare took place in Sunspot Region 9077, 17°N 03°E, which was near the center of the solar disk when the flare occurred. The April flare was in Region 9415, 21°S 47°E, which was more toward to the solar limb (*Spaceweatherlive*, 2014). The spectra of the two flares when the wavelength of 97.7 nm peaked (i.e., 1021 UT and 1918 UT for the July flare and the April flare, respectively) are shown in Figure 3.2. The blue solid and dashed lines show the spectra for the July flare and the April flare respectively when the flux at a wavelength of 97.7 nm reached its peak intensity. The orange solid and dashed lines show the spectra 30 minutes before for the July and April flares, respectively. Both of the flares produced similar enhancements in the X-ray wavelengths ($\sim$ 0.1 to $\sim$ 25 nm) and the EUV wavelengths from $\sim$ 80 to $\sim$ 120 nm, which was different than the flare studied by *Qian et al.* (2010), which was an X17 flare. It was known that center-to-limb variations (CLV) of active regions exist in observation of the solar disk as the sun rotates (*Worden et al.*, 2001). Contrary to the expectation that the EUV irradiance of a limb flare would be lower than a center flare with identical flare classes, the April flare produced slightly higher EUV irradiance. This was possibly because the April flare produced significantly more EUV irradiance in the flare region, so that even after the absorption by the solar atmosphere due to the center-to-limb effect, the EUV emissions still were higher than the center July flare.

In order to explore the difference in the ionospheric response between two solar flares with nearly identical classes, the spectrum of the July flare was grafted to the period when the April flare occurred, and termed the imaginary April flare. The grafted time period was five hours, starting at the flare onset. The same process was done to the April flare to make an imaginary July flare. By moving the flares, it was expected that the ionospheric differences caused by the different flare spectra could be isolated and evaluated. The "imaginary" flares are indicated as dashed lines in Figure 3.1.

The real July X5.7 flare lasted for ∼40 minutes (1003 UT - 1043 UT) (*SWPC*, 2013). The main flare was followed by a small flare (M3.7) that occurred from 1344 UT to 1400 UT (*SWPC*, 2013), about 4 hours later than the main flare. This small flare was also grafted to the April time period as it occurred within 5 hours of the main onset. The real X5.6 April flare lasted for only ∼ 21 minutes (1910 UT-1931 UT) (*SWPC*, 2013). Three small flares (an M3.1 from 0200 UT to 0311 UT, an M8.4 from 0837 UT to 0954 UT and an M5.1 from 1657 UT to 1814 UT respectively) occurred one day before the main flare (*SWPC*, 2013). Since they were before the main flare, these small flares were not grafted to the July time period.

Taking the energy of the solar irradiance during the flare by integrating the irradiances in all wavelengths from 0.1 nm to 175 nm for an hour starting immediately after the solar flare onset, and using the time preceding the flare as a baseline, the real July flare increased the solar irradiance energy by 15.8%, while the real April flare increased the total energy by 11.3%. The ratio between the extra energy in the real July flare and the real April flare was 1.4. This shows that two flares, of the almost identical peak intensity, may differ significantly in total energy. X-rays and EUV are the primary sources of ionization in the E and F regions respectively (*Qian et al.*, 2010; *Leonovich et al.*, 2010), therefore EUV is expected to be a more important source of TEC enhancements and the energy ratio in EUV wavelengths was expected to be more relevant to the ratio of the TEC enhancements caused by the two flares. Comparing the flare energy by integrating the irradiances in EUV wavelengths from 25 nm to 120 nm shows that the real July flare increased energy by 10.98% and the real April flare increased by 7.79%. The ratio between the extra EUV energy in the real July flare and the real April flare was 1.3. This indicates that the differences between the two flares in the total integrated irradiances in X-ray wavelengths were greater than the total integrated irradiances in the EUV wavelengths.

Each flare event was simulated twice, once with the flare, using FISM data and

once without the flare, using a running-12-hour-box-medianed FISM data. The simulations used the solar wind and interplanetary magnetic field data measured by the Advanced Composition Explorer (ACE) and delayed for an appropriate amount of time, and the hemispheric power measured by the National Oceanic and Atmospheric Administration (NOAA) satellites, as shown in Figure 4.1. The solar wind velocity Vx, was unavailable after the July flare onset, so it was replaced with fixed values. All of the external drivers (i.e., IMF, hemispheric power...) except the solar irradiance were the same in the background and flare simulations. The difference between the simulations with and without the flare quantified the ionospheric response to only the solar flare, excluding perturbations due to the other drivers. The perturbation (in percentage) was defined as

$$Perturbation = 100\% \times \frac{Value_w - Value_{wo}}{Value_{wo}},$$

where $Value_w$ and $Value_{wo}$ were the simulated quantity of interest with the FISM spectrum and with the smoothed FISM spectrum respectively. Four regions of the ionosphere were explored to determine the difference between direct and indirect effects of the solar flares: (a) dayside (solar zenith angle, or SZA $< 30\,\mathrm{deg}$); (b) nightside (SZA $> 150\,\mathrm{deg}$); (c) north polar region (or N.P.: latitudes $> 45\,\mathrm{deg}$); and (d) south polar region (or S.P.: latitudes $< -45\,\mathrm{deg}$).

3.4 Results

3.4.1 TEC Perturbation

Figure 3.4 shows the regionally averaged TEC perturbation on July 14-15, 2000. The real and imaginary July flares show similar perturbation profiles. The ratios between the real and imaginary peak ionospheric perturbations were 1.42 on the dayside, 1.37 on the nightside, 1.37 in the north polar and 1.33 in the south polar

regions. The dayside ratio was close to the ratio of the total integrated flare energy between the real and imaginary flares (1.40), and the ratios on the nightside, the north polar region and the south polar region were between the ratios of the total integrated flare energy and of the EUV integrated energy (1.30). As the TEC was decreasing from its most perturbed state, there was a small secondary intensification around three hours after the initial increase, most noticeable on the dayside for both the real and imaginary flares, and the north polar region for the real flare. For the real July flare, this was most likely due to the second intensification in the solar EUV which occurred at this time. For the imaginary flare, a small perturbation on the dayside still occurred, which implied that dynamics played a role in creating this secondary peak.

The nightside TEC perturbation started ∼12 hours later than the perturbations in the other regions. This is because the TEC perturbation did not propagate with the sound speed as the neutral perturbation does (*Pawlowski and Ridley*, 2008). Instead, the flare caused a vertical shift of the ionosphere near the subsolar point, which remained in the same longitudinal sector until recombination caused the perturbation to decay. The nightside only registered an increase when the longitude sector that was near the subsolar point at the time of the flare rotated into this region. The TEC perturbation in this longitude sector was able to last longer than 12 hours, but less than 24 hours, as indicated by the lack of increase in TEC on the dayside 24 hours later. As the north pole was directed more towards to the sun, being the summer hemisphere, the TEC perturbation in the north polar region was larger than that in the south polar region. The dayside was expected to have the greatest perturbation of all the regions; however, that was not the case. The reason why the percentage change in the north polar region was ∼3% higher than on the dayside was that the background TEC on the dayside was greater than that in the north polar region, so the larger real perturbation on the dayside was a smaller percentage change than

what occurred in the northern hemisphere.

Figure 3.5 shows the regionally averaged TEC perturbation on April 6-7, 2001. The real and the imaginary flares once again caused very similar profiles of TEC perturbation. However, the perturbation caused by the imaginary flare was larger than the perturbation caused by the real flare since the imaginary flare contained more energy. As the flare occurred near equinox, the north polar and the south polar regions received similar amounts of flare radiance, and they therefore responded with similar perturbation amplitudes. The secondary flare of the imaginary April flare caused a slight increase in the perturbation on the dayside, north polar and south polar regions; however, these responses were not as sharp as those in Figure 3.4. This indicates that there was something different that occurred during the time periods, even though the flare spectra of the real July flare and the imaginary April flare were identical. The nightside perturbation in response to the main flare was roughly 12 hours later than the response in the other three regions. This is once again because of the lack of propagation in the ionospheric response and the persistence of the perturbation for the 12 hours it took to rotate to the nightside.

The background mid- and low- latitude distribution of TEC at different times before, during and after the July flare are shown in Figure 3.6. Additionally, the TEC difference distributions between the real (middle) and imaginary (bottom) flares and the background simulations are shown. The dark diamond (triangle) indicates where noon (midnight) was located at the time of the plot, while the light diamond indicates the subsolar location at the time of the flare. The black line near the 0 deg latitude marks the geomagnetic equator. The TEC distribution of the non-flare simulation shows two bands of high TEC existing along the geomagnetic equator due to the equatorial dynamo effect. A dayside eastward electric field along with the north component of the geomagnetic field causes ions to drift upward. Forced by gravity and gradients in pressure, ions flow down field-lines away from the magnetic equator

on both sides, resulting in two bands of high TEC along the geomagnetic equator. For the TEC difference, the figures at 1000 UT show the background ionosphere before the flares occurred, which were the same for the real and imaginary flares. This is not zero everywhere because the non-flare simulation was a median filter of the FISM inputs, which means that there were small differences in the EUV drivers between the runs even in non-flare time periods. At 1100 UT, there was an increase in TEC across the entire dayside as seen in the difference plots. At 1500 UT, the perturbed region was mostly confined to just off the magnetic equator, with some perturbations on the nightside in the northern hemisphere. The largest perturbations at 1500 UT were just a bit west of the location of the subsolar point at the time of the flare. From 1500 UT to 2500 UT (i.e., 0100 UT on July 15, 2000), the distribution of the TEC perturbation remained roughly the same but decayed gradually. The region of increased TEC near the longitudinal sector of the subsolar point at the time that the flare occurred lasted longer than 12 hours, such that when it rotated onto the nightside (midnight is indicated by the triangle), a perturbation was registered in this region. The distribution of the TEC perturbations caused by the real and imaginary July flares were very similar. The perturbation was greater in the summer hemisphere (the northern hemisphere) than the winter hemisphere (the southern hemisphere). The TEC perturbations decayed with a similar structure near the geomagnetic equator. However, the perturbation caused by the real July solar flare was stronger than that caused by the imaginary flare. This was because the real July flare contained more energy than the imaginary flare.

Figure 3.7 shows the low- and mid-latitude distribution of the background TEC and the TEC difference before and after the flare on April 6-7 in the same format as Figure 3.6. The first column, 1900 UT, shows the TEC and difference just before the flare. As shown in Figure 3.6, the non-flare simulation shows two bands of high TEC existed just off the geomagnetic equator due to the equatorial dynamo effect. At

2000 UT, the difference plots (middle and bottom) showed that TEC was perturbed across the dayside as seen in the difference plots. At 2400 UT, the perturbation was more concentrated around the geomagnetic equator. At 2900 UT (i.e., 0500 UT on April 7, 2001), the perturbation area remained in the same longitude sector that had rotated to the nightside. Simultaneously, a hole occurred right at the magnetic equator surrounded by areas of increased TEC, which may have resulted from the equatorial dynamo. At 3400 UT (i.e., 1000 UT on April 7, 2001), the perturbation still existed, but was reduced compared to the previous time.

The distribution of the TEC perturbations caused by the real and the imaginary April flares were quite similar to each other. Since the imaginary April flare contained more energy, the perturbation caused by this flare was stronger than the perturbation caused by the real flare. The perturbation areas in the northern and southern hemisphere were similar as it was near equinox when the solar flares occurred. As the TEC returned to a background level, the perturbations were structured by the geomagnetic field. Two bands of high TEC existed at low-latitudes, and bands with decreased TEC existed at mid-latitudes at 2900 UT. These ionospheric structures were caused by the flare-induced neutral wind that flowed away from the flare subsolar region towards the high latitudes. The neutral wind pushed the two high bands away from the geomagnetic equator (as can be seen by the widening of the TEC enhancement bands from 2400 UT to 2900 UT), and as well as pushed the ions further down the field lines around ± 30 deg latitude, resulting in bands of decreased TEC at middle latitudes.

In order to explore how long the ionospheric perturbation at the subsolar point lasted, the average ionospheric density in the region that was within 30° of the subsolar point at the time of the flare was plotted (i.e., the region around the light diamonds in Figure 3.6). Figure 3.8 shows plots of the electron density as a function of altitude and UT in this region above the Earth. The left plots in Figure 3.8 show

the difference in electron density for the July flare. Both the actual and imaginary flare irradiance caused ionization deep into the lower thermosphere with the peak enhancement occurring between 200-300 km altitude. The electron density perturbation below 300 km lasted for around one hour, while the perturbation in the upper ionosphere lasted for approximately 21 hours, which was long enough for that spot to rotate to the nightside and cause a perturbation in TEC. As mentioned above, the real July flare, with more energy, caused greater and longer perturbations.

The right plots in Figure 3.8 show the average electron perturbation within 30° of the subsolar point when the April flare occurred (i.e., the region around the light diamonds in Figure 3.7). The imaginary April flare caused a larger electron perturbation as it contained more energy. However, the electron perturbations caused by the two flares lasted equally long (approximately 21 hours). The slight increase in the density observed above 500 km around hours 24-25 in the imaginary flare may have resulted from the small secondary flare. In addition, the small electron perturbations during hours 0-12 were caused by the three small flares at the beginning of the day before. Examination of the background electron density shows that the structure of the electron density perturbation is similar to the structure of the background electron density. The flares just added a small increment to the background structure.

3.4.2 Background influence

When this study was started, there was an expectation that the ionospheric response to the real July flare and the imaginary April flare (i.e., July flare moved to April) would be similar, and the ionospheric response to the real April flare and imaginary July flare (i.e., April flare moved to July) would be similar. In other words, it was expected that the solar irradiance spectrum would be the dominant controlling factor, while the response of the ionosphere between the real and imaginary flares during the same period would be different. This did not happen. The ionosphere re-

acted quite differently to the exact same flare spectrum in April versus July. Because of this, it was theorized that the background condition had a significant influence on the ionospheric reaction to the flare, while the spectrum had less of an influence. In order to test this, more simulations were conducted. Specifically, the magnetic field topology and season were altered in order to explore their influence. The topology was investigated because the two flares occurred when different magnetic geometries were facing the sun, while the two flares took place in different seasons, so this was also investigated.

3.4.2.1 Geomagnetic field

The real July flare was simulated again, this time using an ideal dipole field. Once again, two simulations were run, one with the 1-minute FISM data and one with the 12-hour-median-filtered FISM data. These were differenced as before, to calculate the perturbation TEC.

Figure 3.9 shows the comparison between the regionally averaged TEC perturbation caused by the real July flare with the IGRF and the July flare with the dipole magnetic field. The solid line represents the TEC perturbation with the realistic geomagnetic field (the same as Figure 3.4), while the dashed line shows the simulation with an ideal dipole magnetic field. Compared with IGRF, the run with the dipole field had a lower average TEC perturbation on the dayside and nightside, with an increased perturbation in the south polar region. In the north polar region, the response was almost identical.

On the dayside, the initial perturbation was almost identical, while the reaction 2-5 hours after the flare differed by a small amount. This difference between the two simulations may have been caused by two mechanisms: first, the smoother topology of the ideal dipole field allowed neutral winds to drag ions more easily away from the dayside along field lines, which led to a more rapid decrease of the dayside TEC

perturbation in the run with the ideal dipole field. Second, the change in field strength influenced the electrodynamics on the dayside, which affected the TEC response in this region. However, it was difficult to determine which factor was more important. The percent perturbation of the averaged TEC in the south polar region of the dipole case was higher than that in the IGRF case. This was because the baseline run with the smoothed FISM drivers and the dipole field had a lower average TEC in the south polar region than the baseline run with IGRF. This caused the percentage variation to be larger in dipole case even though the total perturbation was similar between the two cases.

The perturbations on the nightside showed more disagreement between the two cases. This was because it took 12 hours for the perturbation to rotate to the nightside. During this time, background processes influenced by the structure of the magnetic field, such as the momentum coupling between the ions and neutrals, changed the evolution of the TEC perturbation. This is illustrated in Figure 3.10, which shows the distribution of the mid-latitude TEC difference caused by the real July flare with IGRF (upper row) and with the dipole field (lower row). The magnitude of the TEC perturbations were almost identical in the two cases, while the distribution of TEC in the two cases showed significant differences. The distribution of the IGRF TEC perturbation was along the curve of the magnetic equator, but because the dipole equator was the same as the geographic equator, the TEC perturbation in the dipole case was more symmetric around the geographic equator. The structure of the geomagnetic field was important in determining the post-flare evolution of TEC distribution. As is shown at 2000 UT in Figure 3.10, the perturbation evolved differently as it rotated to the nightside. There are some interesting similarities too. For example, at 1500 UT, between 120°-240° longitude, the perturbation in the north was stronger (although in a different shape). Between 240°-360° longitude, the perturbation in the south was stronger, but the shape was different. This illustrates that the background wind

pattern was similar, but the ion-neutral coupling was different due to the magnetic field topology.

The difference between simulations with and without IGRF was very small compared to the difference between the real July and the imaginary April flares as well as the real April and the imaginary April flares. This indicates that magnetic field structure probably has a large influence on the nightside reaction to the flare, but less of an effect on the other regions.

3.4.2.2 Season

To explore the effect of different seasons on the ionospheric reaction to solar flares, the July 13-15 flare time period (including EUV, solar wind, IMF and hemispheric power inputs) was shifted to March 13-15. In other words, all the forcing of the July run and the March run were the same except they were in different seasons. Figure 3.11 shows the regionally averaged TEC perturbations caused by the July flare (solid line) and the July flare moved to the March time (dashed line). The north polar TEC had a larger perturbation than the south polar TEC in the July case as it was summer in the northern hemisphere, while in the March case, the perturbations in the polar regions were similar because it was near the March equinox. The dayside was only slightly modified by the seasonal difference. The nightside response, however, was significantly different. Rather than a 12-hour delay in the July case, the March nightside perturbation occurred about 5 hours earlier.

Figure 3.12 shows the low- and mid-latitude distribution of the TEC difference caused by the July flare (upper row) and the March flare (lower row). The distribution of the perturbations in the two cases were similar on the dayside (at the diamonds). Although the March flare had a stronger absolute perturbation, as Figure 3.11 shows, the relative perturbations were nearly the same. During the March flare, there was a small perturbation on the nightside. Because this was so fast after the flare onset, the

only mechanism that could have caused this response was a change in the equatorial electroject, which caused a small uplift in the F-region on the nightside, reducing the loss rate, which appeared as a slight density increase in the difference plots. At 1500 UT, the nightside TEC perturbation (near the triangle) was more extensive in March than in July, as is indicated by the large perturbation a few hours after the flare in Figure 3.11. This appears to be primarily due to the fact that the anti-subsolar point was very close to the magnetic field equator during the March time but was far away from the equator during the July time. During July, the anti-subsolar point was about 21.5 degree away from the geomagnetic equator to the south, which led to downward ion flows around midnight. The solar flare intensified this downward flow and caused a decrease in TEC on the nightside, which is shown as the blue region (near the triangle) at 1500 UT. However, in March, the anti-subsolar point was almost right at the magnetic equator, which led to the converging neutral wind primarily pushing the ionosphere upward around midnight. The neutral wind was intensified by the solar flare, causing the TEC to increase on the nightside between 15 and 18 UT. As shown in Figure 3.13, the nightside perturbation in neutral density at 413 km and the perturbation in TEC were well matched in time, although the TEC perturbation reduced more rapidly than the neutral density perturbation.

3.4.3 Comparison with GPS

In order to determine whether the simulated response of the ionosphere to the solar flares was realistic, a comparison between the modeled TEC and measured TEC was conducted. Figure 3.14 shows the relationship between the TEC difference (simulation with the flare minus simulation without the flare) and the SZA at 1100 UT on July 14, 2000. There appears to be two linear relationships, one below 80° and one above 80°. The largest TEC enhancement of approximately 5 TECU existed at SZAs between 10°-15° and 30°-40°. At each solar zenith angle, there is a large

spread in the TEC, which is most likely due to the magnetic field topology causing different ion-neutral coupling processes at the same SZA but very different latitudes and longitudes. The similar plot of a comparison between TEC enhancements derived from TEC observations and their corresponding solar zenith angles during the same flare was present in *Zhang et al.* (2002). The GPS TEC enhancements related to the enhancement in the solar irradiation due to the flare were derived from each temporal TEC curve by removing the influence of the background solar disk irradiation. Near the shadow boundary region with SZAs between 80° and 105°, TEC enhancement still existed, but decreased rapidly with SZA. These findings agree with the GPS observations during the flare on 28 October 2003 obtained in *Zhang and Xiao* (2005b).

Figure 3.15 shows the modeled (top two rows) and measured (bottom row) low- and mid-latitude TEC changes that took place during and after the flare on July 2000. The TEC changes here were derived by subtracting the TEC (model results and measurements) 30 minutes before from the TEC at the plotted time. The top plots show the TEC changes from 1000 UT to 2500 UT on July 14, 2000 in the same format as Figure 3.6, although the differencing method was different. At 1000 UT, small TEC enhancements occurred on the dayside and some regions around the equator on the nightside. At 1100 UT, large TEC enhancements occurred across the dayside and large TEC decreases occurred on the nightside. The narrow positive enhancement that extended across all latitudes in each plot is the morning terminator. The dusk terminator is more difficult to locate. At 1500 UT, the enhancement region on the dayside was moving westward with the subsolar point. From 1500 UT to 2500 UT the enhancement region on the dayside extended slightly in longitude. The middle and bottom plots show the TEC changes from the model and from the GPS observation at all the available GPS sites (*Madrigal*, 2014). At 1000 UT, both the model and GPS showed large areas of negative changes except a small region near dawn. At 1100 UT, both the model and GPS showed the dayside TEC enhancements caused by the

flare near the 300° longitude sector (i.e., the north and south American sector). The model also showed that the enhancement extended to 0°-120° longitude northward of the subsolar point, while the GPS showed decreased values in the same area. At 1500 UT, both model and observation showed enhancements near the dawn sector. The GPS measurements also show some increases near 100° longitude, while the model enhancements in these regions were not as large as those in the observations. At 2000 UT, an enhancement region occurred near 150° longitude in the southern hemisphere and a decrease occurred near -45° longitude in the northern hemisphere both in model and in the observations. At approximately 240° longitude in the model, a decrease was observed, while an enhancement was observed in the data. At 2500 UT, both the model and observation captured the TEC enhancement on the dayside from ~ 100 deg to ~ 240 deg and the depletions at night.

3.5 Summary and Conclusions

In this study, the ionospheric perturbation caused by the two solar flares on July 14, 2000 and April 6, 2001 were examined with the Global Ionosphere and Thermosphere Model. Ideal experiments were conducted by exchanging the two flare spectra to produce imaginary solar flares. An unexpected result was that the ionospheric reaction to different flares with the same background conditions was very similar in temporal and spatial profiles but different in magnitudes, while moving the same flare to a different time caused the ionospheric reaction to be different. This indicates that the background conditions, as well as the total amount of energy in the flare (i.e., the strength and the duration of the flare), are two of the most important parameters in determining the ionospheric reaction to a flare. While the flare spectrum has been shown to be important (*Qian et al.*, 2010), this study indicates that the TEC perturbation is most sensitive to the background state and the total energy input during the flare. When different configurations of the magnetic field were explored, the dif-

ferences were not very dramatic, leading to the conclusion that, while the magnetic field topology is important, changing the topology doesn't appear to significantly alter the ionospheric reaction to the flare, except on the nightside. Conversely, having the flare take place in a different season causes the reaction to the flare to be quite different everywhere but close to the subsolar point. In the polar regions, the perturbation is controlled by the tilt of the Earth, controlling how much flare irradiance enters the atmosphere in the given hemisphere. During equinox, the reactions in the northern and southern hemisphere are almost identical, while the summer hemisphere has a stronger reaction than the winter hemisphere during solstice, as expected. On the nightside, the dynamics are much more complex, and the TEC appears to be controlled by the background winds. Having different wind patterns during different seasons changes the forcing of the ionosphere along field lines, which can dramatically alter the ionosphere over the 12 hours that it takes the perturbation to rotate from the subsolar region to the midnight region. The simulations also show relatively good agreement with GPS observations, when the relationship between TEC enhancements due to the July 14, 2000 flare and local solar zenith angles were explored. The global maps of the TEC enhancements were generally in agreement between the simulation and GPS observations.

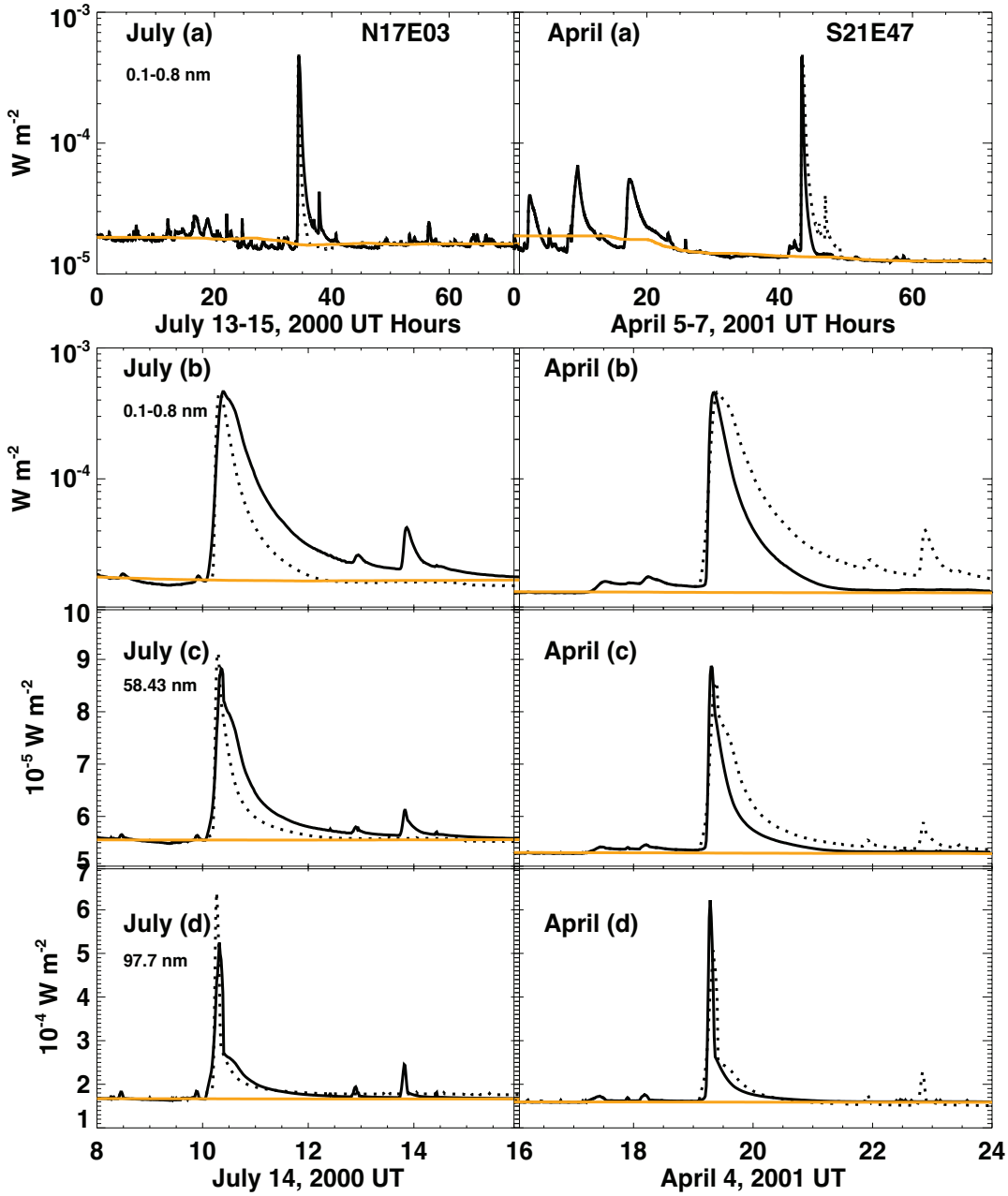

Figure 3.1: (a) Solar irradiances at 0.1-0.8 nm wavelength of the real flare (solid line) and the imaginary flare superposed on the real flare time (dashed line) for 3 days; (b,c,d) flare irradiances at 0.1-0.8 nm, 58.43 nm and 97.7 nm for eight hours. The left and right colomns show the July and April flares respectively.The red lines represent the 12-hour-medianed solar irridiances at corresponding wavelengths.

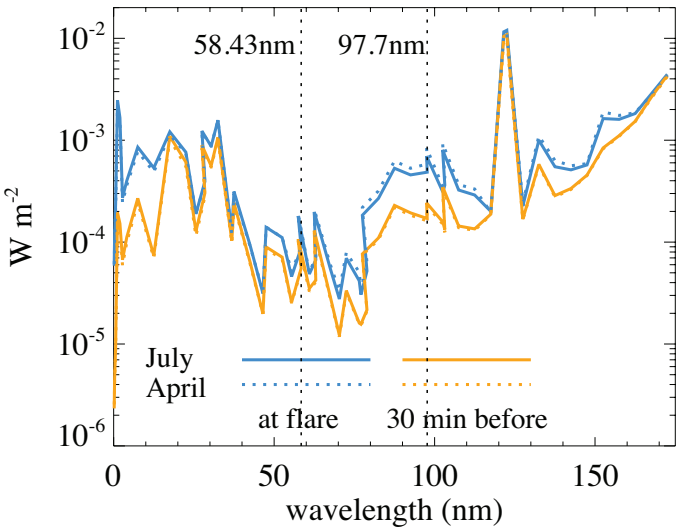

Figure 3.2: Spectrum of solar irradiances when the wavelength 97.7 nm reached peak (1021 UT and 1918 UT for the July and April flares respectively). The blue lines show the spectrum at the peak; the red lines show the spectrum 30 minutes before. The solid lines indicate the July flare; the dashed lines indicate the April flare. The two black dashed lines mark the two wavelengths, 58.43 nm and 97.7 nm.

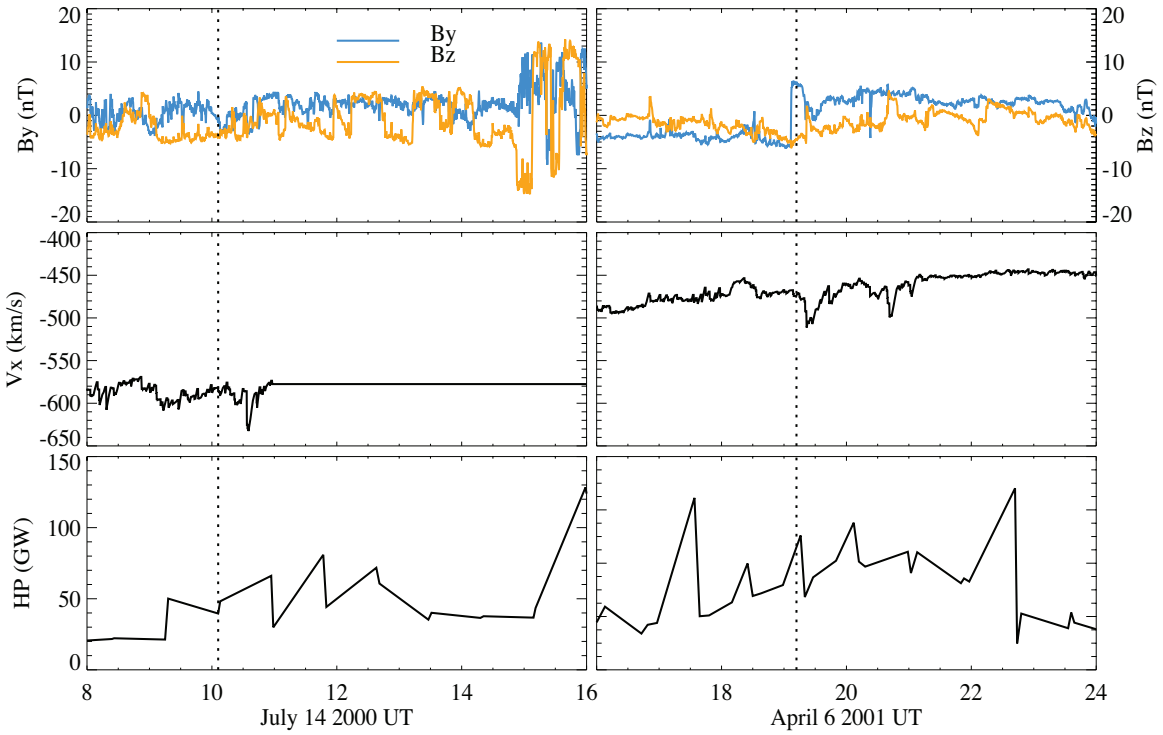

Figure 3.3: The interplanetary magnetic field (top), By (yellow) and Bz (red), solar wind speed (middle), and the northern hemispheric power (bottom) during the July flare (left) and April flare (right). The vertical lines show the onset of the flares.

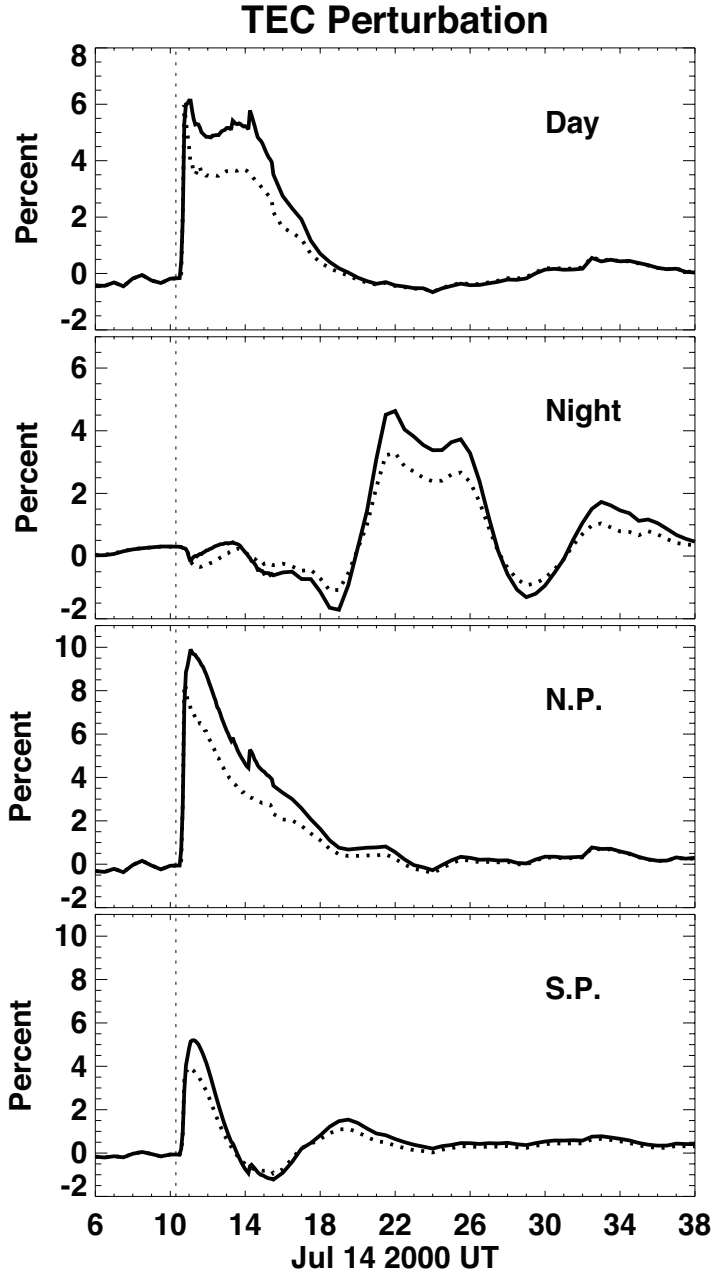

Figure 3.4: Regionally averaged TEC perturbation on July 14-15, 2000 (dayside: solar zenith angle (SZA) < 30 deg; nightside: $SZA > 150$ deg; north polar region (N.P.): $latitudes > 45$ deg; south polar region (S.P.): $latitudes < -45$ deg). The flares began around 1003 UT (indicated as the vertical lines on each panel). The solid line and the dashed line are for the real and the imaginary flares respectively.

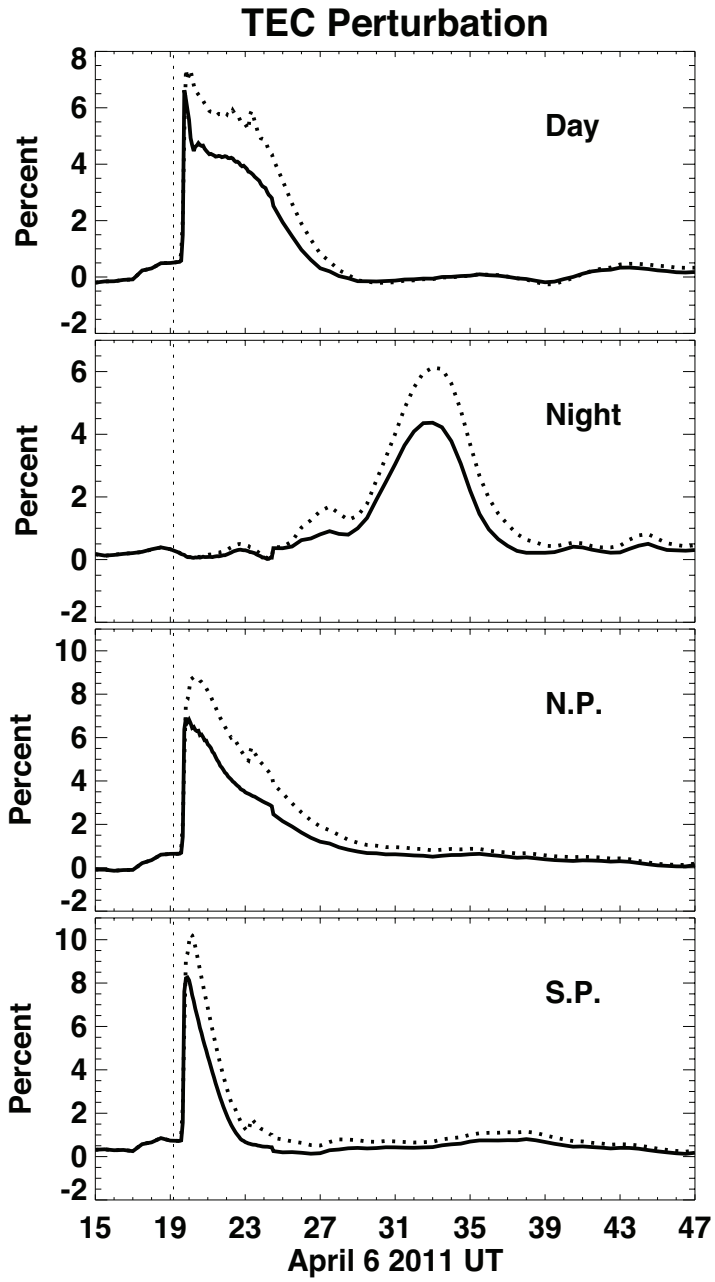

Figure 3.5: Regionally averaged TEC perturbation in April 6-7, 2001. Regions were defined the same as that in Figure 3.4. The flares began around 1910 UT (indicated as the vertical lines on each panel). The solid line and the dashed line are for the real and the imaginary flares respectively.

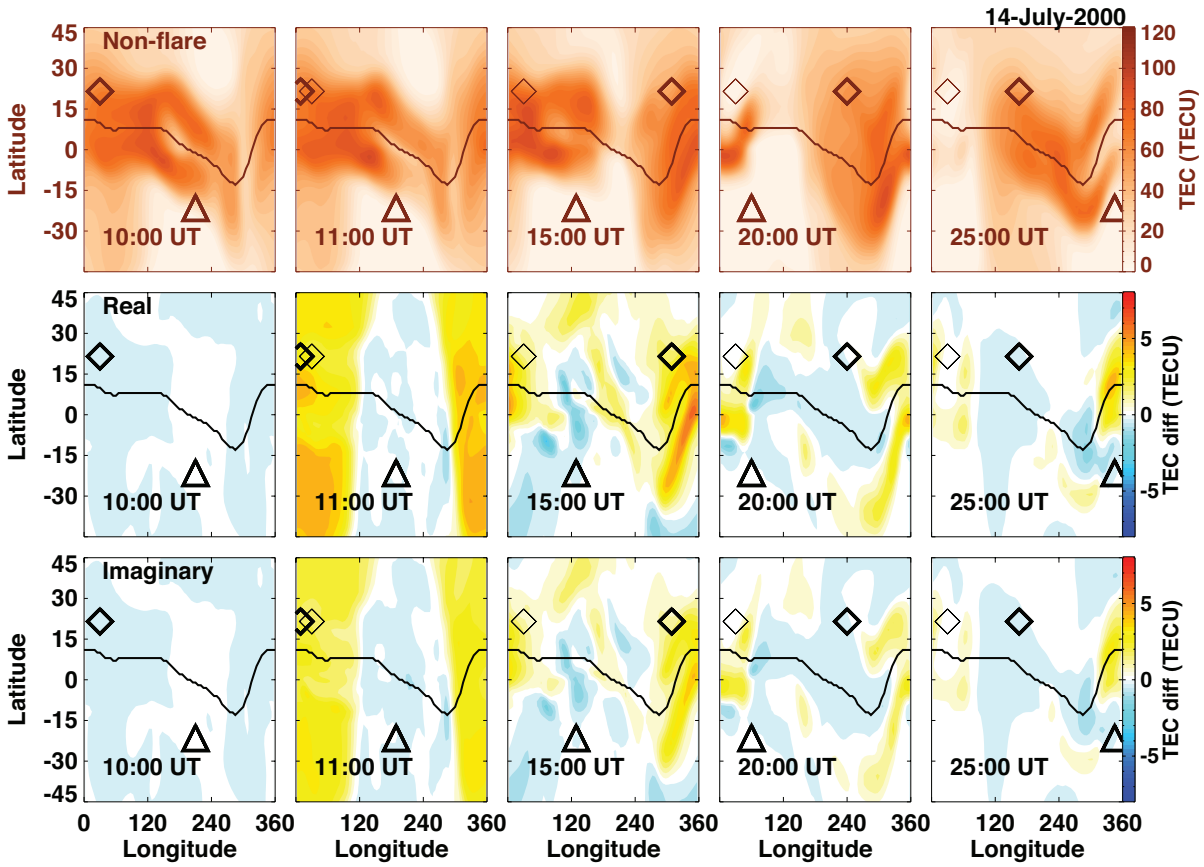

Figure 3.6: The top panels show the mid-latitude TEC of the non-flare simulation on July 14-15, 2000. The mid and bottom panels show the mid-latitude TEC difference at the same time points. The diamonds and triangles mark the local noon and the local midnight respectively. The lighter diamond shows the subsolar point when the flare occurred. The line around 0 deg *Lat* roughly presents the geomagnetic equator. The upper row and the lower row are for the real and the imaginary July flares respectively.

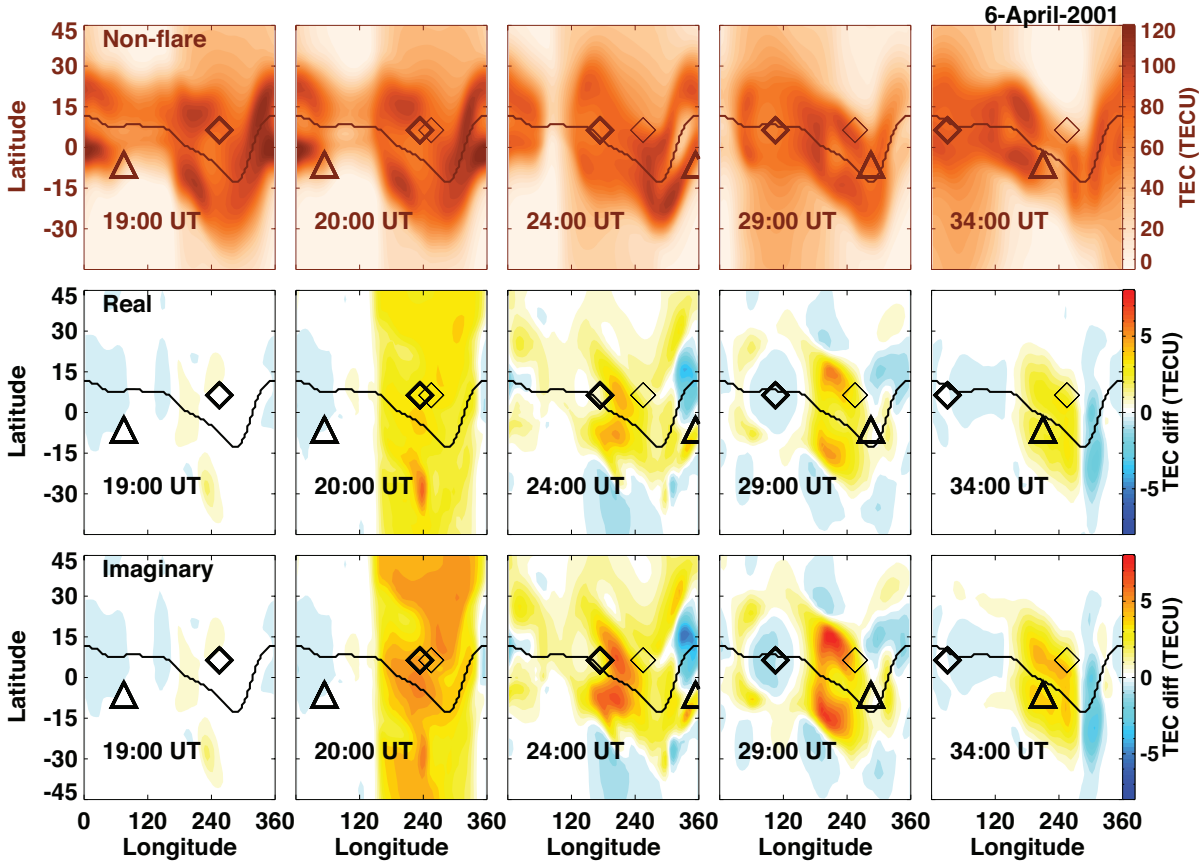

Figure 3.7: The top panels show the mid-latitude TEC of the non-flare simulation on April 6-7, 2001. The middle and bottom panels show the mid-latitude TEC difference at the same time points. The diamonds and the triangles mark the local noon and the local midnight respectively. The lighter diamond shows the subsolar point when the flare occurred. The line around 0 deg *Lat* roughly presents the geomagnetic equator. The upper row and the lower row are for the real and the imaginary April flares respectively.

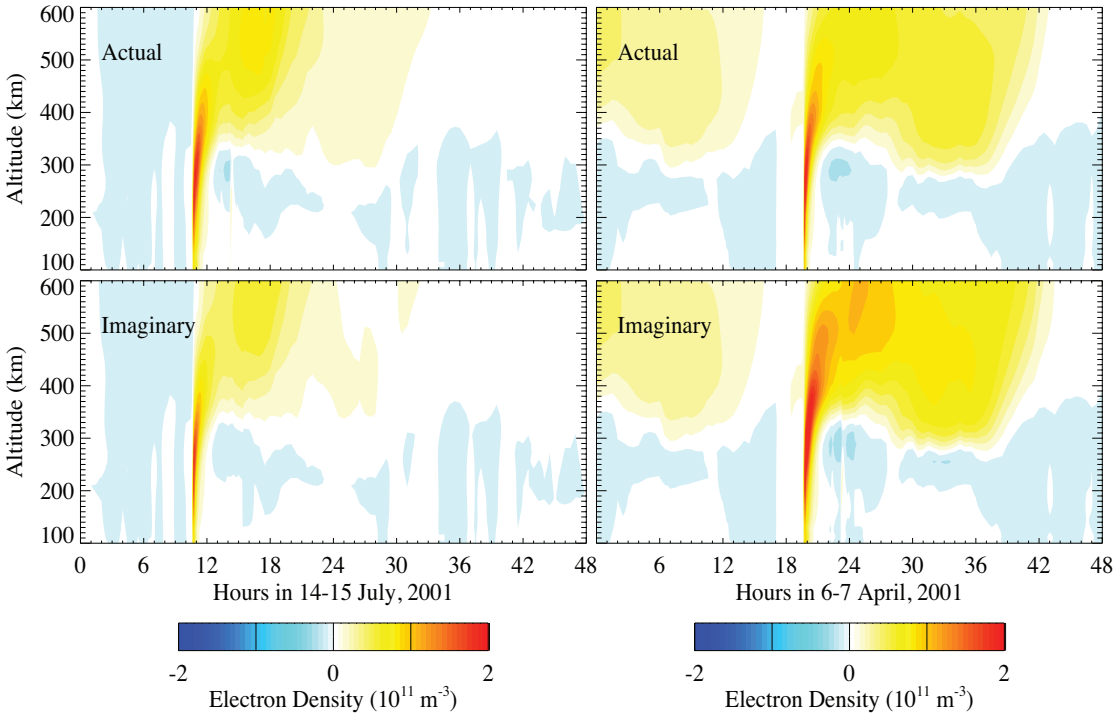

Figure 3.8: Average electron density difference within 30 deg SZA of a fixed site (sub-solar at the flare time) on July 14-15, 2000 (left panels) and on April 04-06, 2001 (right panels). The upper and lower panels present the real and imaginary flare respectively.

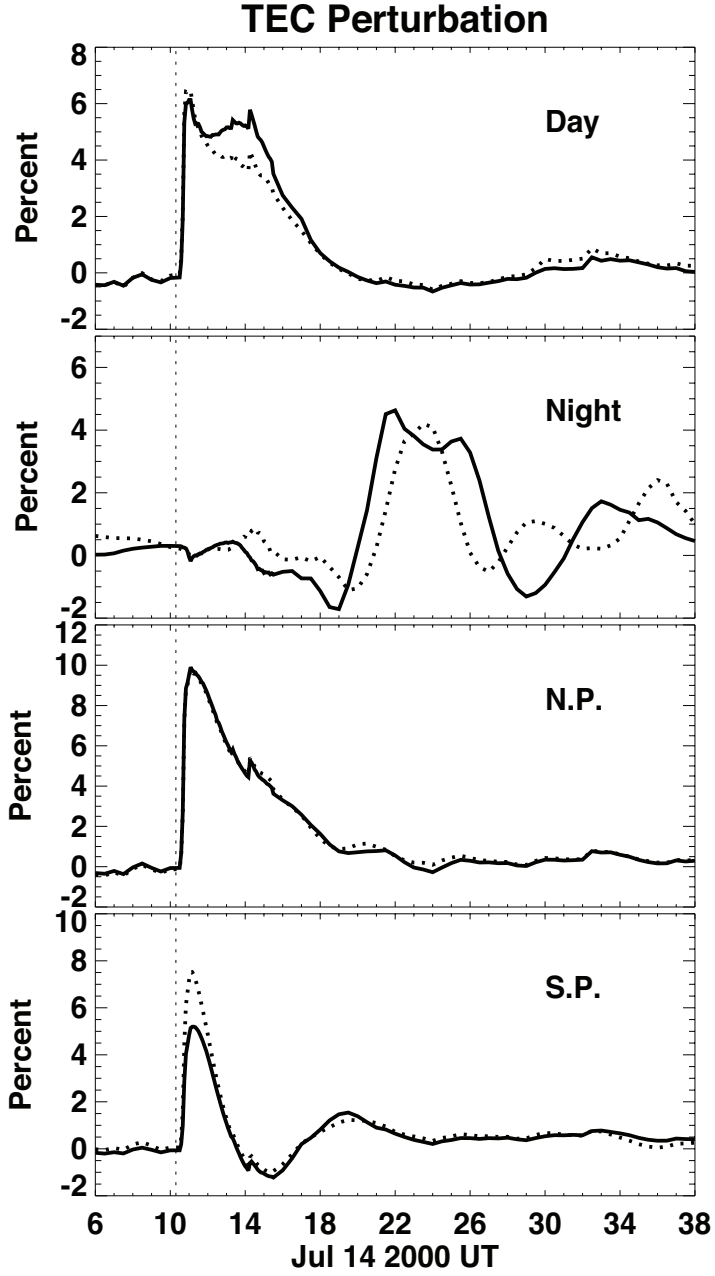

Figure 3.9: Regionally averaged TEC perturbation on July 14-15, 2000. Regions were defined the same as that in Figure 3.4. The solid line and the dashed line are for the real July flare with the realistic geomagnetic field (IGRF) and with an ideal dipole magnetic field respectively. The dashed vertical line indicate the time that the flare onset.

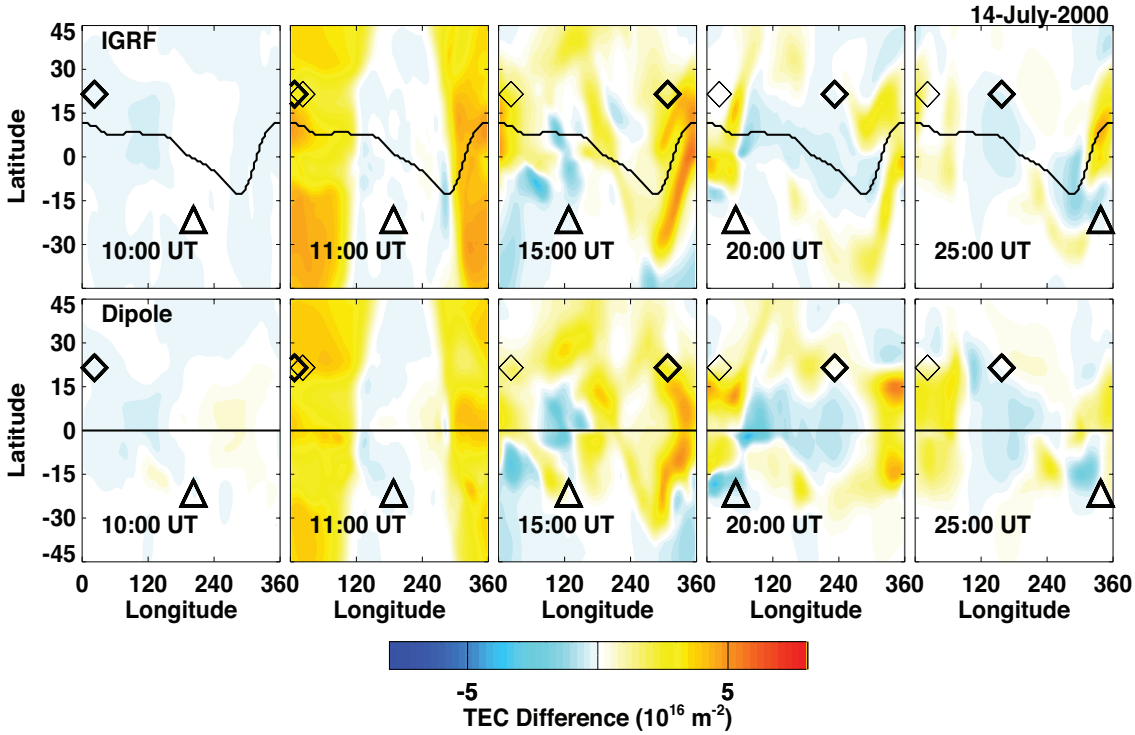

Figure 3.10: Mid-latitude TEC difference on July 14-15, 2000. The diamonds and triangles mark the local noon and the local midnight respectively. The lighter diamond shows the subsolar point when the flare occurred. The line around 0 deg *Lat* roughly presents the geomagnetic equator. The upper row is the real July flare with the APEX; the lower row is the real July flare with dipole field.

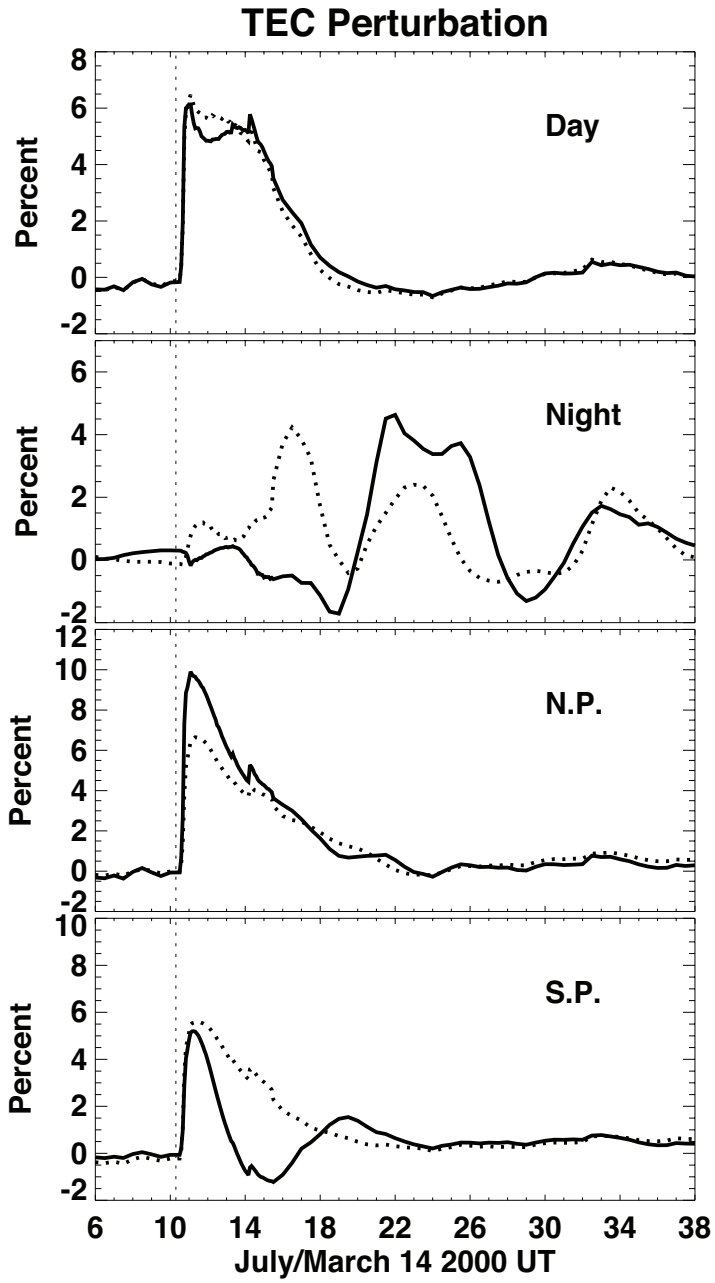

Figure 3.11: Regional average TEC perturbation on July/March 14-15. The solid line and the dashed line are for the July and March flares respectively. The vertical dashed line indicates the time that the flare occurred.

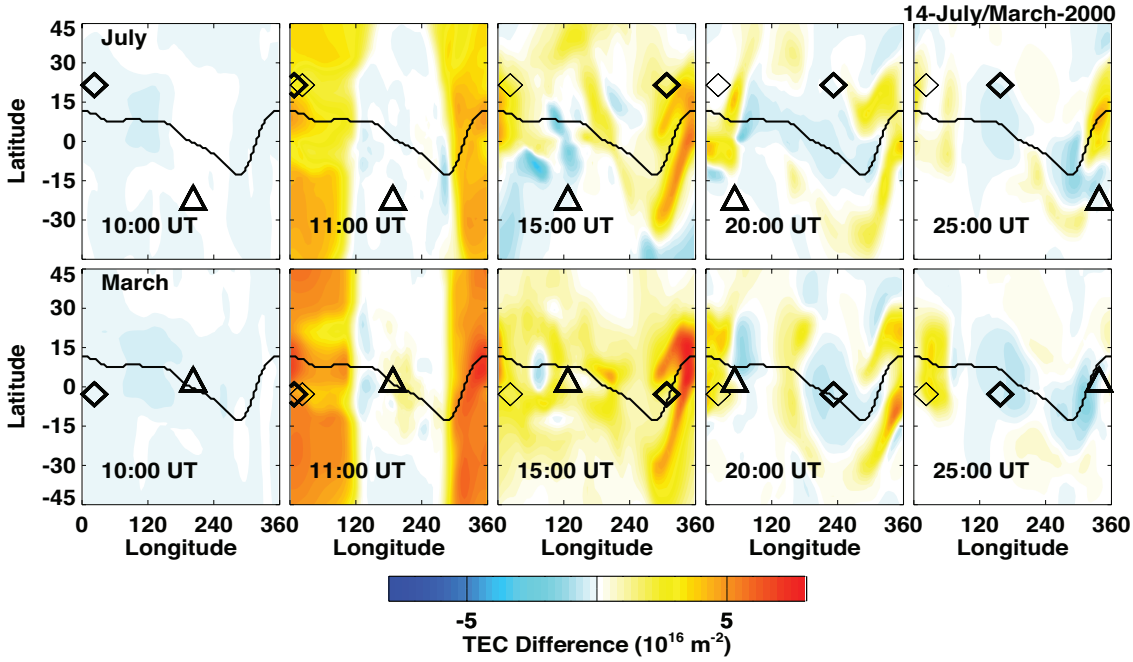

Figure 3.12: Mid-latitude TEC difference on July/March 14, 2000. The diamonds and triangles mark the local noon and the local midnight respectively. The lighter diamond shows the subsolar point when the flare occurred. The line around 0 deg *Lat* roughly presents the geomagnetic equator. The upper row and the lower low are for the July and March flares respectively.

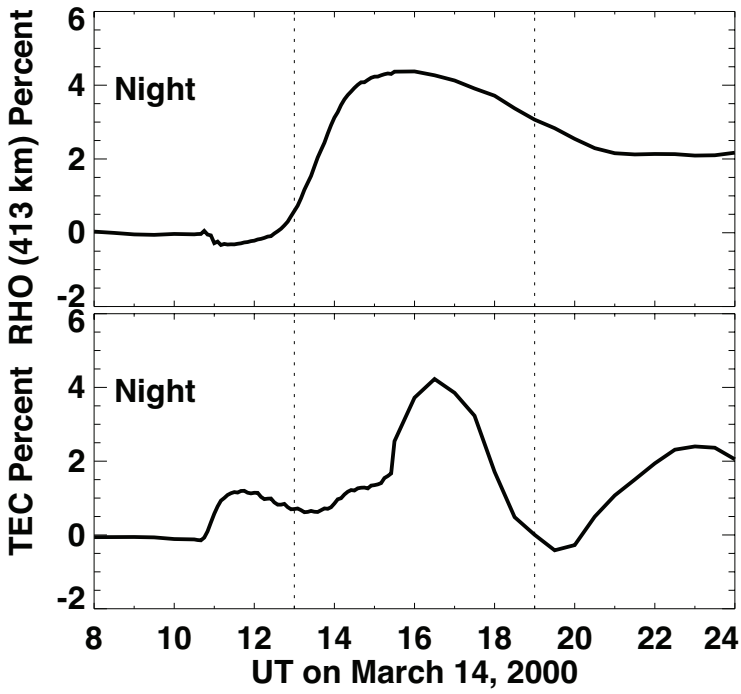

Figure 3.13: Nightside neutral density perturbation at 413-km altitude (top) and nightside TEC perturbation (bottom) on March 14, 2000

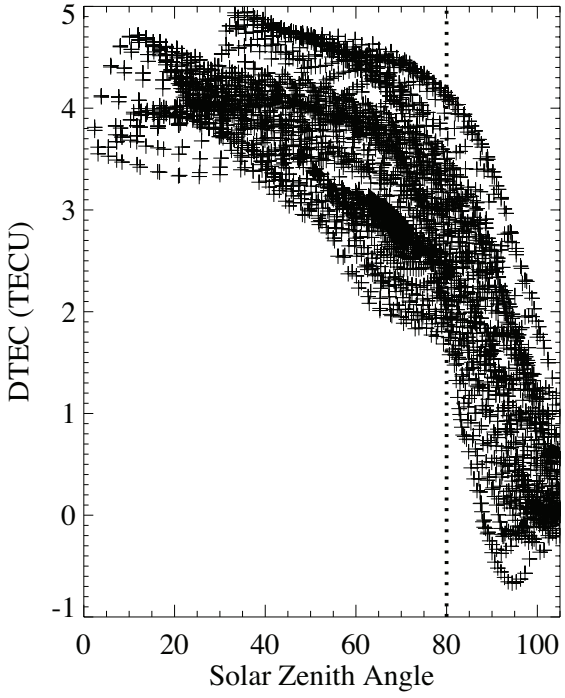

Figure 3.14: TEC difference versus the solar zenith angle on the dayside and the sunlit boundary region at 1100 UT on July 14, 2000. The vertical dashed line marks the SZA of 80°

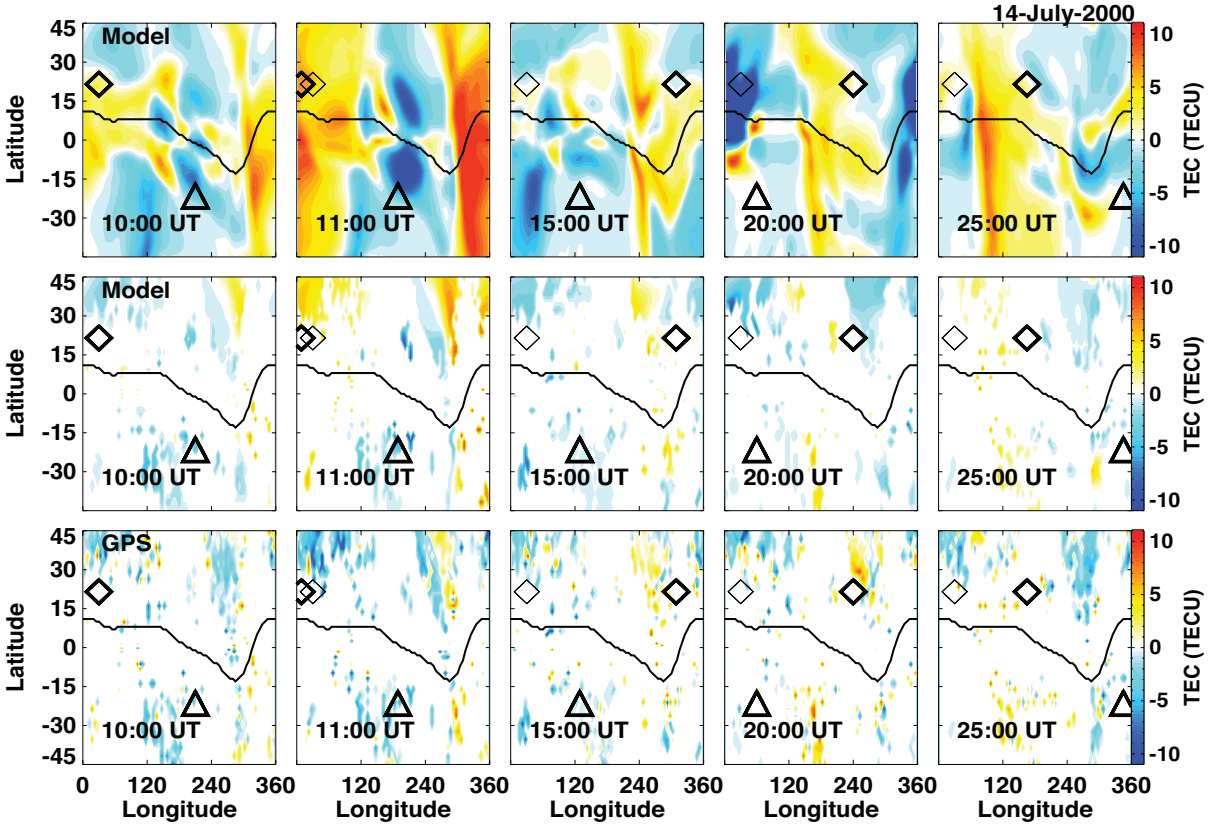

Figure 3.15: The top plots show the TEC changes in model from 1000 UT to 2500 UT on July 14, 2000; the middle (bottom) panels show TEC changes in model (by GPS) at all available GPS sites at the same times as the top plots.

CHAPTER IV

Modeling subsolar thermospheric waves during a solar flare and penetration electric fields

Thermospheric waves occurring around the time of the July 14, 2000 solar flare were investigated using GITM. The simulation results showed that extensive acoustic and gravity waves were excited by the solar flare in the subsolar region. The subsolar buoyancy period at 400-km altitude was approximately 16 minutes. Gravity waves with frequencies lower than the buoyancy frequency traveled from the dayside to the nightside and converged in the longitudinal region that was the anti-subsolar region when the flare occurred. Acoustic waves with frequencies well above the buoyancy frequency propagated upward from approximately 130-km altitude with increasing amplitudes. The power spectra of the vertical neutral winds in the acoustic branch peaked at a period of approximately 13 minutes, just below the buoyancy period. The gradient in pressure was the driver of the two waves, while the ion drag caused a phase delay between the variations in the pressure gradient and the vertical velocity in the acoustic waves. An anti-correlation in the high-frequency component of the vertical neutral wind exists between the subsolar and anti-subsolar points at times away from the flare, which was driven by the rapid variations of the ion flows due to the penetration electric field. It is suggested that the penetration of the high-latitude IMF electric field to low latitudes can drive neutral waves in the equatorial region

through momentum coupling with rapidly changing ion flows.

4.1 Introduction

The ionospheric response to solar flares and penetration electric fields have been widely studied for decades (*Garriott et al.*, 1967; *Davies*, 1990). A solar flare is a sudden enhancement in solar emissions in the X-ray and Extreme-ultraviolet (EUV) wavelengths. These emissions can cause a rapid increase in TEC on the dayside of the ionosphere (*Mendillo et al.*, 1974), and may also result in sudden ionospheric disturbances (SIDs) (*Donnelly*, 1967; *Jones*, 1971; *Stonehocker*, 1970). The recent use of the GPS network in mapping TEC (*Coster and Komjathy*, 2008; *Rideout and Coster*, 2006) has further improved the understanding of the ionospheric response to solar flares. Compared to the extensive studies of the ionospheric response to these rapid variations, the thermospheric response has received less attention. This is partly because the dense neutral atmosphere was expected to react sluggishly to such short time-scale phenomena. However, observations by CHAMP (Challenging Minisatellite Payload) show rapid increases in the neutral density within a few minutes of the flare onset (*Liu et al.*, 2007a). Day-to-night propagating gravity waves, launched by dayside heating during a solar flare, were simulated by a model (*Pawlowski and Ridley*, 2008), but there have been very few studies that have shown whether other neutral waves are generated by solar flares and if so, what their characteristics may be.

Penetration electric fields are another rapid driver of the ionosphere. Two primary electric fields exist in the low- and mid-latitude ionosphere: the wind dynamo electric field and the penetration electric field. The neutral wind dynamo is the interaction between the neutral winds and the electrically conducting ionosphere to generate electric fields and currents (*Kato*, 1956; *KATO*, 1957; *Tarpley*, 1970; *Richmond*, 1989). Penetration electric fields from high-latitudes down to low-latitudes have two primary

sources: the interplanetary electric field and the electric field perturbations associated with substorms (*Kikuchi et al.*, 2000; *Sastri et al.*, 2001). The penetration of the dayside high-latitude electric field, caused by direct connection between the magnetosphere and the IMF, to low-latitudes was first suggested by *Nishida* (1968) and has been extensively studied using the incoherent scatter radars (*Kelley et al.*, 1979; *Fejer et al.*, 1979; *Huang et al.*, 2005). The electric field is confined in the high latitudes by the shielding effect of the Region-2 field-aligned currents (*Vasyliunas*, 1972). However, the slow response of the Region-2 currents to the rapid changes in the Region-1 currents caused by variations in the IMF cannot effectively cancel out the primary high-latitude electric field, resulting in the penetration of electric fields to low latitudes (*Vasyliunas*, 1972; *Southwood*, 1977). The equatorial ExB drift caused by the penetration of the high-latitude electric field during storm time has been simulated by models and the comparison of the vertical ion drifts between model and observation at the magnetic equator has been investigated (*Maruyama et al.*, 2005; *Wang et al.*, 2008). Although the ionospheric response to the penetration electric field has been established, the influence of the penetration electric field on the neutral wind has not received as much attention.

If the effects of viscosity and coriolis force are ignored, rapid changes in the thermosphere have the ability to drive two types of wave structures, acoustic waves and gravity waves (*Gossard and Hooke*, 1975; *Yeh and Liu*, 1974a). The presence of gravity gives rise to a restoring force known as the buoyancy force, which is responsible for the characteristic oscillations at the buoyancy (Brunt-Väisälä) frequency (*Väisälä*, 1932; *Brund*, 1927). In a barometric atmosphere where the pressure (or density) of the air changes with altitude (*Kundu and Cohen*, 2008), the buoyancy (angular) frequency N is defined as:

$$N = \sqrt{-\frac{g}{\rho_0}\frac{d\bar{\rho}}{dz}}, \tag{4.1}$$

where g is the acceleration due to gravity, ρ_0 is the neutral density at a reference altitude and $\bar{\rho}$ is the neutral density in hydrostatic balance. When the pressure (or density) exponentially decreases with altitude, the buoyancy frequency can be further written as:

$$N = \sqrt{\frac{(\gamma - 1)g^2}{{a_s}^2}}, \tag{4.2}$$

where γ is the adiabatic index and a_s is the local sound speed (*Gombosi*, 1998). The local sound speed is given by:

$$a_s = \sqrt{\frac{\gamma p_0}{\rho_0}}, \tag{4.3}$$

where p_0 and ρ_0 are the background pressure and density. Two branches of propagational waves exist in the acoustic-gravity waves. The low-frequency branch in the region where $\omega < \omega_1$ is the gravity wave branch. Here, ω is the angular frequency of the waves and $\omega_1 = Nsin\theta$, where θ is the propagation angle with respect to the vertical direction. The high-frequency branch is the acoustic wave branch where $\omega > \omega_2$. ω_2 is known as the acoustic cutoff frequency and is given by (*Gombosi*, 1998):

$$\omega_2 = N\sqrt{\frac{\gamma^2}{4(\gamma - 1)}}. \tag{4.4}$$

This cutoff frequency is found to be $1.1N$ assuming γ =1.4, resulting in a frequency slightly greater than the buoyancy frequency. The region $\omega_1 < \omega < \omega_2$ is the "cutoff" region where no propagational waves exist. In addition to the presence of the "cutoff" frequency of the acoustic waves, the gravity waves that can not propagate entirely vertically as transverse waves also contribute to the characteristics of the region. The "cutoff" zone is highly propagation-angle dependent, which changes from [0, $N\sqrt{\frac{\gamma^2}{4(\gamma-1)}}$] for a vertically propagating wave to [N, $N\sqrt{\frac{\gamma^2}{4(\gamma-1)}}$] for a horizontally propagating wave.

This study investigated the thermospheric waves around the time of the solar flare on July 14, 2000. The following questions were addressed: (1) Can solar flares and penetration electric fields generate waves in the subsolar thermosphere? (2) How do solar flares and penetration electric fields generate neutral waves? and (3) What are the characteristic of these neutral waves?

4.2 Model

GITM was used to investigate the wave activity on July 14, 2000. The horizontal grid resolution of the simulations was set to 5° longitude by 2.5° latitude. IGRF was specified in the simulation. The flare spectrum calculated by FISM (*Chamberlin et al.*, 2007) was used as the input for the solar irradiance. The high-latitude electric field was specified by the Weimer potential model (*Weimer*, 2005), which was driven by measured IMFs and solar wind velocities from the ACE satellite. The ionospheric potential was updated every 60 seconds in the simulation, including both the high-latitude and dynamo potentials. In this case, the dynamo boundary was put at 65° magnetic latitude, which allowed a bit of the high-latitude potential to "leak" to the low latitudes. This altered the high-latitude boundary potential on the dynamo, resulting in penetration electric fields.

Figure 4.1 shows the interplanetary magnetic field and solar wind data measured by the Advanced Composition Explorer (ACE) and the hemispheric power measured by the NOAA satellites. The IMF and solar wind measurements were delayed for an appropriate amount of time using a simple ballistic propagation method (*Ridley*, 2000). The magnetic field was relatively quiet when the flare occurred at approximately 1003 UT. The IMF then had large rapid variations from approximately 1430 UT-1700 UT, with B_z oscillating between +/- 13 nT. The solar wind velocity data were unavailable shortly after the flare, so a fixed value was used for the remainder of the day.

4.3 Methodology

This study explored the wave activity in the thermosphere around the time of the X5.7 solar flare on July 14, 2000. The flare event was simulated twice, once accounting for the flare by using the solar irradiance from FISM, and once ignoring the flare by using running 12-hour-box-medianed FISM data. The simulation with the flare represented the thermospheric response during the flare time and was termed the raw simulation, whereas the difference between the two simulations, with and without the flare, describes only the thermospheric response to the flare under given background conditions and is termed the difference simulation.

Figure 4.2 shows the solar irradiance at X-ray wavelengths (0.1-0.2 nm) (top) and at extreme ultra-violet (EUV) wavelengths at 58.43 nm (middle) and 97.9 nm (bottom) on July 14, 2000. The yellow lines represent the 12-hour-medianed irridiances at the corresponding wavelengths. The flare lasted for around 40 minutes (1003 UT - 1043 UT) (*SWPC*, 2013). The peak irradiance at 0.1-0.8 nm classifies the flare as an X5.7. The main flare was followed by a small flare (M3.7) that occurred from 1344 UT to 1400 UT. Two EUV wavelengths, 58.43 nm and 97.7 nm, were plotted because of the large photoabsorption and photoionization cross sections for atomic Oxygen at 58.43 nm, and for molecular Oxygen and Nitrogen near 97.7 nm (*Huffman*, 1969). It was expected that the EUV flux would cause thermospheric responses through photoabsorption of the neutrals and enhanced exothermic chemical reactions due to the increased ionization, as well as through momentum and energy coupling between ions and neutrals.

A Fast Fourier Transform (FFT) was used to analyze the spectra of the wave activity at the subsolar point at 400 km altitude. The buoyancy frequency was calculated to determine the frequency ranges of acoustic and gravity waves. To examine the temporal evolution of the wave activity, the spectra of the raw and differenced vertical wind using a 2-hour moving window were also investigated. The low-frequency

component of the neutral dynamics was obtained by taking the mean value of a running 30-minute window on the simulation data. The high-frequency component was obtained by subtracting the low-frequency component from the simulation data.

4.4 Results

4.4.1 Large-scale gravity waves initiated by a solar flare

Pawlowski and Ridley (2008) found that the intense dayside heating by a solar flare launches nightward propagating gravity waves that transport energy to the nightside at velocities close to the local sound speed plus the bulk neutral wind velocity. This gravity wave that they investigated converged near local midnight after it passed through the dawn and dusk sectors and over both poles, although it was not precisely midnight.

The present study found that this thermospheric disturbance converged at the longitude sector that was midnight when the flare occurred instead of the longitude sector of midnight when the wave arrived approximately 3.5 hours later. Figure 4.3 shows the 5 K isolines from the differenced temperature between the run with and without the flare at 400 km altitude beginning at 1000 UT and ending at 1400 UT on July 14, 2000. The color of the isolines transitions from blue to red with increasing UT. Just after the flare onset at 1030 UT, the initial 5K isoline appeared, approximately covering the dayside half of the globe. As time progressed, the disturbance gradually traveled toward the nightside and eventually converged at the anti-subsolar point when the flare occurred (thick blue triangle). It took approximately 3.5 hours for the disturbance to travel from the dayside and converge at the flare-time midnight. The converging sector was approximately 60° eastward of midnight at the time of convergence (thin blue triangle). One interesting feature was that the isoline at 1345 UT extended a little wider toward the actual midnight, which could possibly be a conse-

quence of the background day-to-night neutral wind pushing the convergence region westward.

Propagation of the global-scale gravity wave caused by the flare was also observed in the vertical neutral wind. The low-frequency component of the vertical neutral winds was obtained by averaging the differenced simulation data with a sliding window of 30 minutes. This was important since there existed a clear high-frequency component due to acoustic waves, which will be discussed in the subsequent section. The differenced low-frequency vertical winds caused by the flare are shown in Figure 4.4. The thick diamond and thick triangle represent the subsolar and anti-subsolar sectors when the flare occurred, and the thin diamond and thin triangle mark the locations of the subsolar and anti-subsolar regions at the time of the plot. The color scale is saturated at ± 0.6 m/s to show the nightside perturbation more clearly. At 0930 UT, prior to the flare onset, the vertical neutral wind was almost unperturbed. At 1030 UT, the flare caused an increase of over 0.6 m/s in the low-frequency vertical neutral wind on the dayside and within the north polar region. The nightside and south polar region were not disturbed by the flare at this time. The difference in the response to the flare between the north and south polar region was attributed to seasonal variation. Because it was summer in the northern hemisphere, the south pole did not receive as much solar irradiance and was not able to respond immediately. At 1130 UT, the dayside perturbation was weakened. Some downward flow existed in the north polar region. At 1230 UT, the low-frequency vertical flow across the entire dayside and north polar region was downward, while the sectors near dawn, dusk, as well as the nightside started showing upward flows. At 1330 UT, downward flow still existed on the dayside and in the north polar region, and the region of upward flow had moved toward the anti-subsolar sector at the flare onset (dark triangle) with an increasing amplitude. This indicates that a possible large-scale gravity wave in the vertical neutral wind was excited by the flare as described by *Pawlowski and Ridley*

(2008). This gravity wave traveled from the dayside to the nightside in approximately 3.5 hours. Once again, instead of converging on the anti-subsolar point at the particular time, the wave converged on the sector that had been the anti-subsolar point when the flare onset occurred.

Figure 4.5 shows the low-frequency component of the differenced vertical winds at the flare-time subsolar (solid line) and anti-subsolar (dotted line) points from 0900 UT to 2400 UT. It took approximately 2 hours for the gravity waves to start influencing the anti-subsolar point. The daytime perturbation had a narrow half "sine" shape with only a very shallow negative phase, while the nighttime perturbation had a wider full "sine" shape. The anti-subsolar wave phase was positive for $\sim$3 hours and turned negative around 1600 UT.

4.4.2 Acoustic waves initiated by a solar flare

The time period from 0930 UT to 1130 UT was chosen to investigate the acoustic waves around the solar flare on July 14, 2000. Figure 4.6 shows the temporal variation of the vertical neutral wind at the subsolar sector (moving on the Earth) from 0930 UT to 1130 UT. The dashed vertical line at 1003 UT marks the time of the flare onset. The subsolar location was chosen because it is the most sensitive to changes in solar irradiance, since the solar irradiance directly travels through the atmosphere with no reduction due to the projection to the vertical direction. Panel (a) shows the attitudinal profiles of the raw vertical neutral wind from 100 km to 800 km. The disturbances propagated from lower altitudes to higher altitudes with increasing amplitudes. Because the group velocity of the acoustic waves has the same sign as the phase velocity in all the three directions (*Yeh and Liu*, 1974a), the energy was transported upward by the acoustic waves, and in the absence of damping, the amplitude of the wave energy was conserved during the upward propagation. The exponential decrease in the neutral mass density causes the amplitudes of the vertical

wind to grow in amplitude as $\rho_0^{-1/2}$ accordingly, where ρ_0 is the basic unperturbed neutral density (*Lindzen*, 1967, 1968, 1981; *Hodges*, 1969). An upward-propagating acoustic wave in the thermosphere, caused by a sudden intense enhancement of high-latitude Joule Heating, was simulated by *Deng et al.* (2008), who pointed out that the magnitude of the vertical wind perturbation also increased with altitudes.

The small disturbances below 300 km are difficult to observe, as a result of the ever-growing amplitude of vertical propagation waves. To accentuate the amplitude of the low-altitude waves, the attitudinal profiles of the raw vertical wind were scaled according to the square root of the ratio of the neutral mass density at the plotted altitude (ρ) and the background density (ρ_{400km}), which gave the normalized vertical wind $VN_{UP.norm}$ of:

$$VN_{UP.norm} = VN_{UP} \cdot \sqrt{\frac{\rho}{\rho_{400km}}}. \tag{4.5}$$

As shown in panel (b) of Figure 4.6, the normalized vertical neutral wind shows the perturbation at low altitudes. The perturbation in the vertical wind started from approximately 130-km altitude during the flare. The rapid enhancement in energy deposition at low altitudes led to heating in the lower and middle thermosphere, which drove an enhancement in the vertical pressure gradient. This disturbed the hydrostatic equilibrium in the low thermosphere, causing acceleration of the vertical wind accordingly. The source of the acoustic waves will be discussed in more detail below.

Panel (c) shows the raw vertical neutral wind at 400-km altitude. The amplitude of the vertical wind peaked at 10 m/s around 1025 UT. The amplitude of the variation was approximately 7 m/s, with a baseline of 2-3 m/s upward flow. This baseline is simply the dayside lifting of the atmosphere due to the slow thermospheric temperature increase throughout the day time. Panels (d) and (e) show the attitudinal profiles of the differenced vertical wind and the normalized differenced vertical

wind respectively. The amplitude of the flare-driven waves increased with altitudes as expected. Based on the positions, P(10.27, 300) and Q(10.35, 600) in panel (a), the vertical propagation speed was calculated as 1042 m/s. It was close to the averaged local sound speed, $\sqrt{\frac{\gamma p}{\rho}}$, of approximately 1074 m/s at 400-km altitude from 0930 UT to 1130 UT in the simulation. Panel (f) shows the differenced vertical wind at 400-km altitude. The flare increased the vertical neutral wind by 6 m/s within 20 minutes. The perturbations of the acoustic waves due to the flare decreased rapidly and were mostly gone in about one hour. Note that this perturbation in the differenced vertical wind was a combination of the large-scale gravity wave and the acoustic waves. Panel (g) shows the low-frequency component of the differenced vertical wind in Panel (f) and that the positive phase of the large-scale gravity wave went through the subsolar point (the same as the solid line from 0930 UT to 1130 UT in Figure 4.5). Panel (h) shows the high-frequency component of the differenced vertical wind (i.e., the acoustic waves). The acoustic wave activity decreased quite quickly after the flare occurred. Both the acoustic waves and the gravity wave had amplitudes of $\sim$3 m/s.

In order to determine whether these variations are similar to the actual neutral wind, the model results were compared qualitatively with measurements. This is quite difficult to do on the dayside, since most Fabry-Perot Interferometers (FPIs) operate on the nightside only. Further, data from the near equatorial region was quite limited during this time period. Therefore, data from the Arecibo FPI (18.4° N 66.6° W) on July 3, 2000 was compared to the simulation on July 14, 2000 at the same location as shown in Figure 5.5. Obviously, an exact match is not expected since these are different days. But, qualitatively, the waves observed in the simulation results (pre-Flare) are comparable to the waves observed by the FPI at Arecibo. The FPI uses the brightness measurements at 630 nm, which gives the wind conditions around 240-250 km (*Friedman and Herrero*, 1982; *Jacka*, 1984; *Hernández*, 1988). Figure 5.5 shows a good agreement in the magnitude of the vertical wind ($\sim$5 m/s) between the

observation and the simulation results.

4.4.3 Wave analysis

Figure 4.8 shows the Fast Fourier Transform (FFT) of the vertical neutral wind at 400-km altitude for the one-day period on July 14, 2000. The blue dashed line marks the average buoyancy period of approximately 16 minutes; the red dashed vertical line marks the period of 13 minutes where the peak power occurred. The top panel shows the FFT of the raw vertical neutral wind, while the bottom panel shows the FFT of the differenced vertical wind.

As previously discussed, the waves with frequencies (periods) that are greater (smaller) than the buoyancy frequency are the acoustic waves, which lie on the left side of the buoyancy period (blue line), while the gravity waves, with frequencies (periods) below (above) the buoyancy frequency, lie on the right side of the buoyancy period. Three peaks existed in the range of the acoustic-wave frequencies, of which the period with the most wave energy was 13.2 minutes. This was also the period in which the differenced vertical wind contained the most energy in the acoustic-wave range. As this peak frequency (0.0013 s^{-1}) was comparable to the buoyancy frequency (0.0010 s^{-1}), the potential energy associated with the buoyancy force may have had a similar order of magnitude between the kinetic energy and elastic energy of the acoustics (*Yeh and Liu*, 1974a). Therefore, the acoustic waves that contained the most energy may be highly affected by the buoyancy force. The power above the buoyancy period stayed at a relatively low value until a period of about 60 minutes, at which point the power increased to a value comparable to the peak in the acoustic-wave range. This lack of wave activity possibly indicates the existence of the "cutoff" region in the acoustic-gravity waves.

One interesting feature to note is that the magnitudes of the acoustic waves around the period of 13 min and the magnitudes of the gravity waves with periods of ∼200

min were comparable as shown in the spectra of the vertical wind in Figure 4.8. However, the amplitudes of the temperature and density perturbations in the gravity waves were approximately 1-2 orders greater than in the acoustic waves. Figure 4.9 shows the FFT of the raw neutral temperature (top row) and neutral density (bottom row) at 400-km altitude for 24 hours on July 14, 2014. Assuming that the waves with periods smaller than the buoyancy period (blue line) were acoustic waves while the waves above the buoyancy period were gravity waves, the left figures (2 minutes to 30 minutes) show the spectra of the acoustic waves, while the right figures (30 minutes to 24 hours) show the spectra of gravity waves. In the acoustic wave spectra, there existed a clear power peak at 13.3 minutes with temperatures similar to the vertical wind (Figure 4.8). The peak in density was difficult to decipher. The amplitude of the acoustic wave in the temperature was about 0.5 K, while the amplitude of the gravity wave was about 10 K. For the density, the amplitude of the gravity wave was almost two orders larger than the magnitude of the acoustic waves. The theoretical ratio of the perturbed density between the gravity and acoustic waves is $\frac{a_s N^2}{\omega_g g}$, while that of the perturbed pressure between the gravity and acoustic waves is approximately $\frac{N^2}{\omega_g^2}$, where ω_g is the angular frequency of the gravity waves. Assuming the buoyancy frequency N was 10^{-2} rad/s and the angular frequency of the gravity waves was 10^{-3} rad/s, the ratio of the perturbed density and pressure between the gravity wave and acoustic waves were 10 and 100, respectively. Based on a simple ideal gas assumption, the ratio of the perturbed temperature between the gravity and acoustic waves should be 10-100. Note that the ideal gas assumption is not correct for gravity waves, which might introduce some error in the ratio of temperature. These theoretic ratios are in relatively good agreement with the ratios produced by the model.

Figure 4.10 shows how the power spectrum of the raw vertical wind at 400-km altitude at the subsolar point changes as a function of time using a two-hour window that extended one hour before and one hour after the centered time. The top panels

shows the spectra of the raw vertical wind, while the bottom panel shows the spectra of the differenced vertical wind. The frequency range above the buoyancy frequency (red solid line) contains the spectrum of acoustic waves, and the range below shows that of gravity waves. Before the flare occurred, a small amount of acoustic and gravity wave activity occurred. Most of the acoustic wave energy occurred at a frequency of 0.002 s^{-1} (i.e., a period of 8 minutes). It can be inferred that gravity waves propagated with varying directions as a function of time before the flare onset, since the range of the forbidden zone appeared to change with time.

There was essentially no wave activity in the differenced vertical neutral wind, because there were essentially no differences in the drivers before the onset of the flare. The flare occurred at 1003 UT as marked by the red dashed vertical line. The perturbation started one hour ahead in the two-hour spectrum because each window was chosen one hour before and after the centered time. Extensive acoustic waves were excited by the flare, with a power peak slightly above the buoyancy frequency. Acoustic waves with high frequencies that were above 0.004 s^{-1} (i.e., periods less than 4 minute) existed during the flare. The power of the acoustic waves peaked right above the buoyancy frequency. The "cutoff" zone almost disappeared when the flare occurred. This indicates that the gravity waves that were excited by the flare propagated more horizontally. The "cutoff" region showed up again around 1200 UT, indicating the time when the horizontally-propagating gravity waves died down.

The bottom panel in Figure 4.10 shows the two-hour spectra of the differenced vertical wind at the corresponding point. As discussed above, the flare excited extensive acoustic wave and gravity wave activity. Similar to the raw spectra, the power of the acoustic waves in the differenced wind peaked right above the buoyancy frequency. Horizontally-propagating gravity waves were also excited at the flare onset and died down in about 2 hours. The wave activity due to the second much smaller flare, from 1400 UT to 1500 UT can be observed in the bottom panel.

Significant wave activity existed in the raw vertical wind from 1500 UT to 2000 UT. These were caused by the IMF disturbances during this time period. The IMF disturbances drove more intense acoustic waves than gravity waves when compared to those driven by the flare. This means that the flare caused both acoustic waves and a broad spectrum of gravity waves while the IMF disturbances were more likely to excite strong acoustic waves and weak gravity waves of a specific frequency at the subsolar point near the equator.

Figure 4.11 shows the temporal variations of the buoyancy frequency at the subsolar point at 400-km altitude for the flare simulation (top) and for the non-flare simulation (bottom), respectively, for July 14, 2000. The percent difference in the buoyancy frequency and the temperature difference between the flare and non-flare simulations are shown in the third and bottom panels, respectively. The flare and non-flare simulations show similar variations prior to the flare onset. The buoyancy frequency decreased by almost 2% two hours after the flare occurred, before it gradually recovered to an undisturbed status. This was because the increase in the solar irradiance enhanced the thermospheric temperature on the dayside when the flare occurred. Under the ideal gas assumption, the sound speed (Eq. (4.3)) can be further expressed as:

$$a_s = \sqrt{\frac{\gamma n k T}{n \bar{m}}} = \sqrt{\frac{\gamma k T}{\bar{m}}}, \tag{4.6}$$

where k is the Boltzmann constant, n is the number density of the neutrals, T is the neutral temperature and $\bar{m}$ is the number-density weighed molecular mass. Inserting a_s into Eq. (4.2), the buoyancy frequency can be written as:

$$N = \sqrt{\frac{\gamma - 1}{\gamma} \frac{\bar{m} g^2}{k T}}. \tag{4.7}$$

Since the buoyancy frequency is inversely proportional to the temperature, it de-

creased when the temperature increased during the solar flare. The influence of the solar flare on the subsolar buoyancy frequency lasted for about 14 hours, which is the same time-scale of the temperature perturbation.

4.4.4 Sources of the wave generation

Figure 4.12 shows the temporal variations of the vertical neutral wind (Panels (a) and (c)) and neutral parcel acceleration in the vertical direction (Panels (b) and (d)) between 125 km and 200 km from 0930 UT to 1130 UT. The region below ∼125 km was not discussed because there were high-frequency waves existing due to a lack of hydrostatic equilibrium in the lower boundary from MSIS. Panels (a) and (b) show the high-frequency variation while Panels (c) and (d) show the low-frequency variation. The high-frequency results were obtained by subtracting the mean value within a running 30-min window from the original data to remove the lower frequency variability in the acceleration results. The low-frequency components were obtained by subtracting a running 4-hour average from the running 30-min average. 30-min and 4-hour were chosen because the periods of acoustic waves (∼13 min) were well below 30 min and the periods of the gravity waves (∼ 2 hours) were between 30 min and 4 hours. Therefore, the acoustic and gravity waves should be observed in the high-frequency and low-frequency results respectively.

The acceleration in the vertical direction a_z was calculated by using:

$$a_z = -\frac{1}{\rho}\frac{\partial p}{\partial z} + g, \tag{4.8}$$

where ρ is the neutral density, p is pressure, z is altitude and g is the local gravitational acceleration. The variations in the high- and low-frequency components in the acceleration were caused by the variations in the pressure gradient.

Assuming the acoustic waves propagate upward with a wave vector k and the gradient in pressure was the dominant force in the wave propagation, the equation of

wave motion in the vertical direction can be expressed as:

$$\rho_0(\frac{\partial \tilde{u}}{\partial t} + u_0\frac{\partial \tilde{u}}{\partial z}) = -\frac{\partial \tilde{p}}{\partial z}, \tag{4.9}$$

where the background neutral density and vertical velocity are ρ_0 and u_0, respectively, and the perturbation in the pressure and in the vertical velocity are $\tilde{p}$ and $\tilde{u}$, respectively. Consider a wave propagation in form of:

$$\tilde{u} = \tilde{u}_0 e^{i(kz-\omega t)}, \tilde{p} = \tilde{p}_0 e^{i(kz-\omega t)}, \tag{4.10}$$

where $\tilde{u}_0$ and $\tilde{p}_0$ are the amplitudes of the vertical velocity and pressure in the acoustic waves. Substituting into Eq. 4.9 gives the relationship between $\tilde{u}$ and $\tilde{p}$:

$$\tilde{p} = \rho_0(\frac{\omega}{k} - u_0)\tilde{u}. \tag{4.11}$$

Consider the dispersion relation of the acoustic waves:

$$\frac{\omega}{k} = a_s, \tag{4.12}$$

where a_s is the local sound speed as shown in Eq. (4.3). Substituting Eq. (4.12) into Eq. (4.11), and accounting for the fact that the vertical motion u_0 was much less than the local sound speed c_s, there is theoretically a positive phase relation between the pressure and vertical velocity. However, the simulation showed no apparent cross correlation (correlation coefficient of -0.16) between the vertical wind and pressure gradient at 200 km altitude in the acoustic waves (i.e., high-frequency results). Conversely, if the ion-drag is also considered as a dominant driver besides the pressure gradient, the wave equation in the vertical direction becomes:

$$\rho_0(\frac{\partial \tilde{u}}{\partial t} + u_0\frac{\partial \tilde{u}}{\partial z}) = -\frac{\partial \tilde{p}}{\partial z} + \rho_i\nu_{in}(u_i - \tilde{u}), \tag{4.13}$$

where ρ_i is the ion mass density, ν_{in} is the ion-neutral collision frequency and u_i is the vertical ion velocity. For simplicity, taking the ion velocity as constant of zero and substituting Eq. (4.10), the relationship between $\tilde{u}$ and $\tilde{p}$ is:

$$\frac{\tilde{p}}{\tilde{u}} = \frac{\rho_0}{k}[(\omega - ku_0) + \mathbf{i}\nu_{in}]. \tag{4.14}$$

Eq. (4.14) can be rewritten as:

$$\frac{\tilde{p}}{\tilde{u}} = \frac{\rho_0}{k}((x + \mathbf{i}y), \tag{4.15}$$

where $x = \omega - ku_0 \approx \omega$, and $y = \nu_{in}$. Eq. (4.15) can further be written as:

$$\frac{\tilde{p}}{\tilde{u}} = \frac{\rho_0}{k}\sqrt{x^2 + y^2}(cos\theta + \mathbf{i}sin\theta) = \frac{\rho_0}{k}\sqrt{x^2 + y^2}e^{\mathbf{i}\theta}, \tag{4.16}$$

where

$$\theta = arccos\frac{x}{\sqrt{x^2 + y^2}}. \tag{4.17}$$

This indicates there was a phase delay of θ between the waves in the vertical velocity and the pressure. Because $sin\theta > 0$ and $cos\theta > 0$, The delayed phase θ was between 0 and $\frac{\pi}{2}$. Taking the ion-neutral collision frequency as 0.06 s^{-1} ($\nu[O^+, O_2]$ at 200 km) (*Schunk and Nagy*, 2009) and the frequency of the acoustic waves as ω =0.02 rad/s, there existed a phase delay of approximately $\frac{7}{18}\pi$. The average buoyancy period at 200 km was approximately 14.5 minutes. Taking the phase delay of $\frac{7}{18}\pi$, a real time delay of 2.82 minutes would be expected between the vertical velocity and the pressure. In the simulation, a maximum correlation of 0.52 was reached at a time delay of 2 minutes. The effect of the ion drag on the vertical neutral motion will be discussed in more detail below.

The vertical wind and the vertical acceleration were anti-correlated in the gravity

waves with a correlation coefficient of -0.67 at 200 km altitude as shown in Panels (c) and (d). The theoretical ratio between the perturbation pressure and the vertical perturbation velocity in the gravity waves in the thermosphere is:

$$\frac{\tilde{p}}{\tilde{u}} = -\frac{\rho_0}{\omega k}(N^2 - \omega^2), \tag{4.18}$$

where ρ_0 is the background neutral density, k is the wave number in the vertical direction such that $k > 0$ when the phase speed is upward (*Kundu and Cohen*, 2008). Because the frequencies of the gravity waves were smaller than the buoyancy frequency, i.e, $\omega^2 < N^2$, the vertical velocity in the gravity waves was inversely proportional to the perturbation in pressure gradient. This anti-correlation was well modeled in the simulation. Therefore, the non-hydrostatic equilibrium due to the perturbation in the pressure gradient was the main source of the gravity waves. In order to better show the large-scale gravity wave activity, the low-frequency components of the vertical wind and vertical acceleration from 0000 UT to 2400 UT are shown in Figure 4.13. The phase speed of the gravity waves was downward as expected. The periods of the gravity waves were approximately 2-3 hours.

The correlation between the perturbation in the vertical motion and the perturbation in the pressure gradient in both the acoustic waves and the gravity waves show good agreement between the simulation results and the theoretical calculations. This leads to an inference that the gradient in pressure was the driver of the two waves, and that the ion drag force caused a phase delay between the pressure gradient variations and the vertical motion in the acoustic waves.

Figure 4.14 shows vertical profiles of the subsolar percentage changes in the neutral temperature, mass density and pressure, as well as the difference in the vertical acceleration with respect to 10:06 UT. The temporal change of the vertical acceleration was caused primarily by the change in the pressure gradient since gravity does not change with time. The six colored lines represent six UTs between the flare onset

and the flare peak. The temperature increased by approximately 0.4% at maximum around 140 km because of the increase in heating due to the flare. The temperature decreased above 200-km altitude. The mass density increased by ∼0.9% around 250 km and decreased both in the regions below 130 km and above 320 km. The causes of these changes in the high-altitude temperature and the mass density are still unclear to us and need further study. The combination of the changes in temperature and mass density led to a net increase in the pressure around 200 km and a decrease in the pressure above ∼300 km. The spatial and temporal variations in the pressure further led to the perturbations in the vertical acceleration of the neutral particles at different altitudes, as is shown in the rightmost panel in Figure 4.14. These perturbations thus started launching the thermospheric gravity and acoustic waves, as described above.

4.4.5 Acoustic waves generated by penetration electric fields

Figure 4.15 shows the high-frequency components of the neutral and ion vertical winds as well as the dynamo electric field at 400 km at the subsolar and anti-subsolar locations. The high-frequency component was obtained by subtracting a running 30-minute average from the simulation data. There exists an anti-correlation between the dayside and nightside flows. This can be quantified through a cross correlation of -0.17 for the neutrals and -0.47 for the ions.

As shown in third panel of Figure 4.15, the subsolar (red line) and the anti-subsolar (blue line) zonal dynamo electric fields were anti-correlated with a cross correlation of -0.75. A dawn-to-dusk electric field existed in the equatorial ionosphere, resulting in an eastward electric field on the dayside near the magnetic equator, and a westward electric field on the nightside. With the combination of the northward magnetic field, the eastward (westward) electric field on the dayside (nightside) caused upward (downward) drifts of the ions. This is shown in the middle panel of Figure 4.15, where the vertical ion velocity is positively correlated with changes in the dynamo electric

field. The plotted ion drifts are the high-frequency component, while the electric field is not filtered. This indicates that the rapid variations in the vertical motion of the ions were controlled by the changes in the dynamo electric field. The dynamo electric field is primarily determined by the divergence of currents driven by neutrals colliding into ions, forcing them to move in the neutral wind direction. This divergence creates field-aligned currents that drives the F-region electric field (*Maeda and Kato*, 1966; *Kelley*, 2009). Further, the low-latitude electric field is driven by the penetration of the magnetospherically-driven high-latitude electric field to lower latitudes (*Nishida*, 1968; *Kelley et al.*, 1979; *Huang et al.*, 2005). These penetration electric fields may change rapidly due to the variability in the high-latitude ion flows.

The changes in the electric field observed in the equatorial region were strongly correlated with the interplanetary electric field (bottom panel of Figure 4.15). The IMF-driven electric field E_{IMF} was calculated based on the formula:

$$E_{IMF} = V_s B_s sin(\theta/2), \tag{4.19}$$

where V_s is the solar wind speed in the sunward direction, $B_s = ({B_y}^2 + {B_z}^2)^{\frac{1}{2}}$, and θ is the angle between the two components of the IMF, B_y and B_z (*Kan and Lee*, 1979). The correlation between low-latitude ion drifts and the interplanetary electric field has been confirmed by observations, which showed the variations in the observed dayside ionospheric electric field at middle and low latitudes in association with the IMF electric field (*Huang et al.*, 2005). Practically in GITM, the high-latitude electric fields (driven by the IMF and solar wind) penetrate to the equator via the high-latitude boundary condition on the dynamo solver, which was placed at 65 deg magnetic latitude, allowing some of the potential to leak to lower latitudes. Since the *Weimer* (2005) empirical electric field model was used in this study, with an update time of 1-minute, any variations in the IMF or solar wind velocity caused changes in the high-latitude potential pattern and the boundary of the dynamo solver.

This resulted in changes in the low-latitude dynamo potential, and therefore the low-latitude electric field. As shown in the third and bottom panels of Figure 4.15, the dynamo electric field at the subsolar point generally had a positive correlation with the IMF-driven electric field (0.7), while the electric field at the anti-subsolar point was negatively correlated with the interplanetary electric field (-0.7). This was because the projection of the eastward interplanetary electric field in the high-latitudes to the equatorial region resulted in an eastward electric field on the dayside and a westward electric field on the nightside.

The high-frequency component of the vertical neutral velocity appeared to be driven by rapid variations in the ion flows which were most likely driven by the rapid variations in the dynamo electric field. The variability of the vertical ion velocity was approximately three times greater than that of the vertical neutral velocity. The high-frequency component of the vertical neutral wind was a consequence of the combination of the vertical ion drag pushing the neutrals up and down, and a restoring force to bring the atmosphere back to a hydrostatic equilibrium. Both gravity and the gradient in pressure can act as a restoring force at the wave frequencies (periods of $\sim$15 minutes) near the buoyancy frequency (*Yeh and Liu*, 1974a).

In order to thoroughly explore the cause of the anti-correlation between the dayside and nightside wave activity, three additional simulations were conducted. First, the high-frequency components in the vertical neutral and ion velocities in the non-flare simulation were explored as shown in Figure 4.16. Without the perturbation due to the flare, the cross correlations were -0.18 for the neutrals and -0.47 for the ions, which were similar to those with the flare. This indicates that the fluctuations in solar irradiance due the flare did not lead to the anti-correlations.

Second, a simulation with the solar flare turned on, but the dynamo turned off was conducted to further investigate the role of the dynamo electric field in producing the anti-correlation. As shown in Figures 4.17, the day-night correlations in the high-

frequency components of the vertical neutral wind were -0.17 with dynamo on (top), 0.01 with dynamo off (middle) and -0.21 in the difference between the dynamo-on and the dynamo-off simulations (bottom). By removing the perturbation due to the flare (bottom), the anti-correlation due to the dynamo electric field was enhanced to -0.21. This indicates that the dynamo electric field, instead of the flare, was the primary factor causing the anti-correlation in the neutrals. Figure 4.18 shows the day-night correlation in the high-frequency component of the vertical ion velocity in the same format as Figure 4.17. The correlation coefficient was positive (0.45) in the dynamo-off case (middle), and was reduced to -0.60 when only the perturbation caused by dynamo existed (bottom). Therefore, the dynamo electric field also was primarily responsible for the anti-correlation in the ions.

Finally, a simulation that included the flare, had the dynamo on and a constant solar wind input of $B_x = 0.0$ nT, $B_y = 0.0$ nT, $Bz = -2.0$ nT and $Vx = 400.0$ km/s was conducted to exclude the possibility that the solar-irradiance-induced dynamo field caused the anti-correlation. As shown in Figure 4.19, the day-night correlations in the high-frequency components of the vertical neutral velocity were -0.17 (top) with the measured IMF input, 0.05 with the constant IMF input (middle), and -0.18 in the difference between the simulations using the measured IMF and the constant IMF inputs (bottom). This means that the variation in the IMF, instead of the variation in the solar irradiance, drove the anti-correlation in the neutrals. Figure 4.20 shows the day-night correlation in the high-frequency components of the vertical ion velocity in the same format as Figure 4.19. The correlation coefficient was 0.35 with the constant IMF input but with the flare (middle), and was decreased to -0.60 when only the dynamo field due to the IMF existed (bottom). This means that the variations in the penetration electric field, rather than the dynamo field induced by the solar irradiance, resulted in the anti-correlation in the ions. In summary, the penetration electric field was primarily responsible for the anti-correlation in the high-frequency

components of the vertical motion between the dayside and nightside atmosphere.

The relatively low correlation between the vertical ion drifts on the dayside and nightside (i.e., -0.47), is most likely due to the dramatic difference in the dayside and nightside conductivity and the structure of the high latitude electric field in the noon and midnight sectors. Further, the magnetic field structure has significant longitudinal variation, so even if the electric field were exactly anti-correlated between the dayside and nightside, the velocity due to ExB drifts would be different. It is argued here that the ion drifts drive neutral vertical motion, which causes high frequency waves to occur. This means that, because the ion motion only has a correlation of -0.47, the neutral motion must have a worse correlation than this. Given that the ion drag force is dependent on the electron density and the neutral mass density, which are very different on the dayside and nightside of the Earth, it is no surprise that the correlation between the dayside and nightside vertical winds is only -0.17.

It should be mentioned that the treatment of the penetration electric field in the model is not sufficiently physical, because the simulations used the IMF as the only source of the high-latitude electric field, and the *Weimer* (2005) model does not allow the electric field to penetrate to low latitudes by default. GITM gets around this by putting an artificial boundary at a fixed location, allowing the potential to penetrate to the equator. This method does not accurately represent the electrodynamics that actually occurs in the magnetosphere-ionosphere system, but allows the physical processes to be approximated. However, the model suggests a mechanism that the neutral waves at low-latitudes can be driven by the high-latitude electric field through coupling with ions when penetration electric fields are present.

4.5 Summary and Conclusion

The non-hydrostatic waves surrounding the July 14, 2000 solar flare were investigated by using the Global Ionosphere Thermosphere Model. The results show that the sudden enhancements in EUV and X-ray fluxes during a solar flare can cause rapid enhancements in the thermospheric temperature at low-altitudes, which can drive acoustic and gravity waves. Gravity waves with frequencies well below the buoyancy frequency appear to travel from the dayside to the nightside (*Pawlowski and Ridley*, 2008), and converge at the longitudinal sector which was the anti-subsolar point when the flare occurred instead of the anti-subsolar point at the time of arrival (about 3.5 hours later). Acoustic waves appear to propagate vertically from the low-altitude source region with ever-increasing amplitudes. The wave spectra of the vertical neutral wind show that the acoustic waves have a power peak at approximately 13 minutes during the flare. The flare caused horizontally-propagating gravity waves that decreased the "cutoff" region, where no propagational waves exist. The flare also caused a decrease in the buoyancy frequency in the subsolar thermosphere of approximately 2%, but this change may not easily be observed in measurements since it is mostly likely smaller than measurement errors. The amplitudes of the acoustic waves and gravity waves were similar in the vertical neutral wind, while the amplitudes of the gravity waves were about 1-2 orders of magnitude greater than the acoustic waves in the neutral density and temperature, as expected. The gradient in pressure was the driver of both the gravity and acoustic waves, while the ion drag caused a phase delay of approximately $\frac{7}{18}\pi$ at 200 km between the variations in the pressure gradient and the vertical velocity in the acoustic waves.

There existed an anti-correlation in the high-frequency component of the vertical neutral wind between the subsolar and anti-subsolar points a couple of hours after the flare occurred. This anti-correlation appears to have been driven by the rapid variations of the ion flows due to the dynamo electric field. Variations in the dynamo

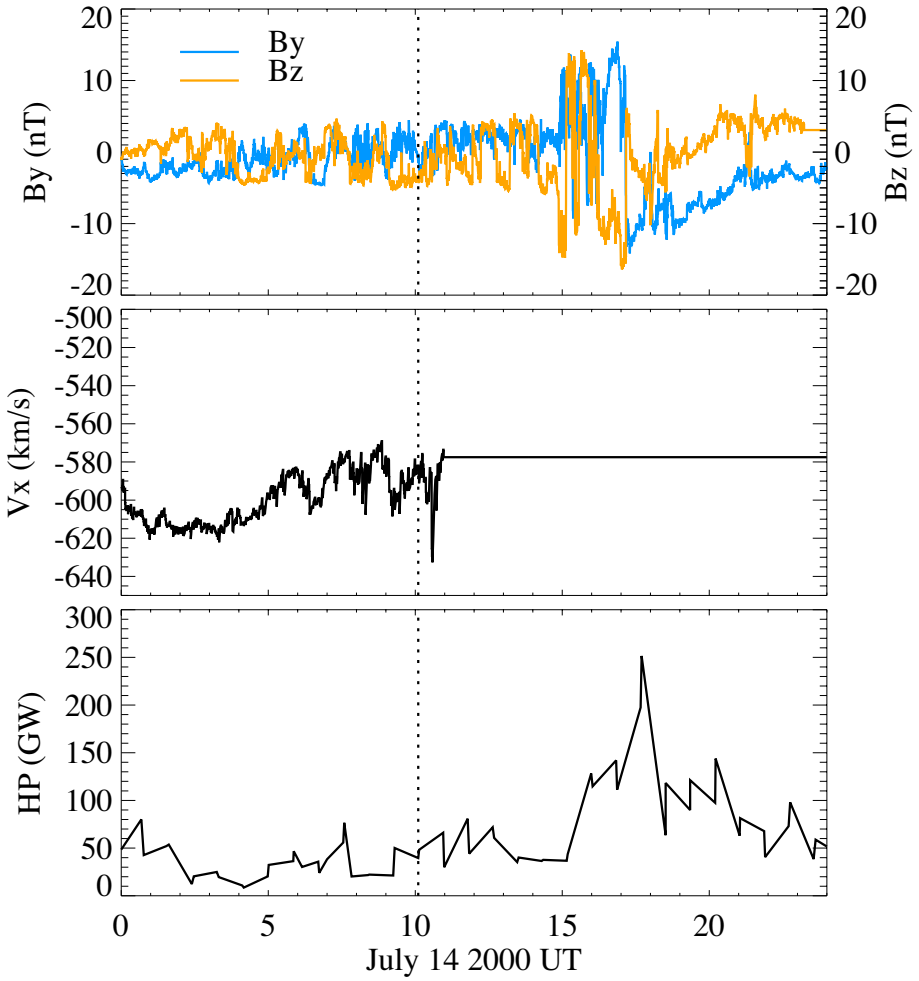

Figure 4.1: The inteplanetary magnetic field By (blue) and Bz (yellow) (top), solar wind speed (middle), and the northern hemispheric power (bottom) during the flare. The vertical dotted line shows the onset of the flare.

electric field were controlled by the penetration of high-latitude electric field, driven by the interplanetary magnetic field, to low latitudes. The penetration of IMF-driven electric field to the equatorial regions resulted in an eastward and a westward electric field on the dayside and nightside, respectively. Neutrals were perturbed by ion drag, then restored to a near-hydrostatic condition, resulting in waves with frequencies that were close to the characteristic neutral buoyancy frequency. During the flare, the anti-correlation in the high-frequency neutral wind between the subsolar and anti-subsolar points was disturbed due to the excitation of acoustic waves on the dayside by the flare.

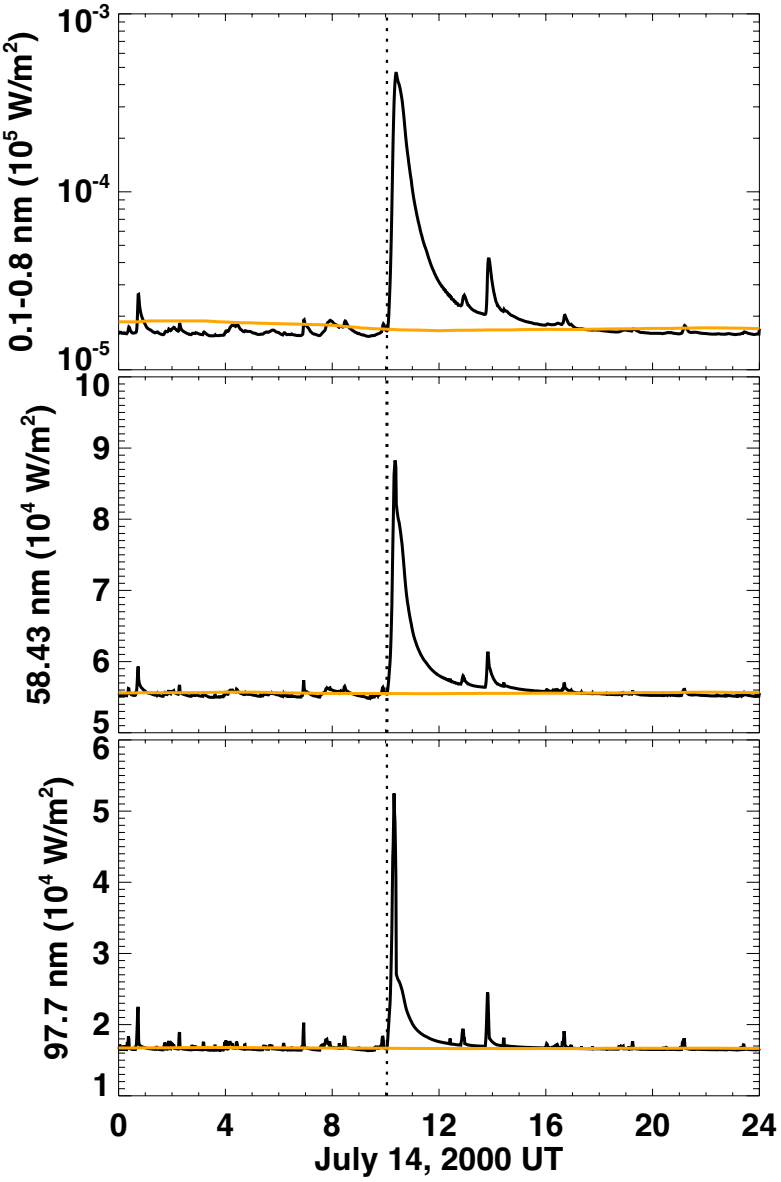

Figure 4.2: Temporal variations of the solar radiance at X-ray wavelengths (0.1-0.8 nm) (top) and at extreme ultra-violet (EUV) wavelengths: 58.43 nm (middle) and 97.7 nm (bottom) on July 14, 2000. The dashed vertical line at 1003 UT marks the flare onset. Yellow lines represent the 12-hour medianed solar irridances at corresponding wavelengths.

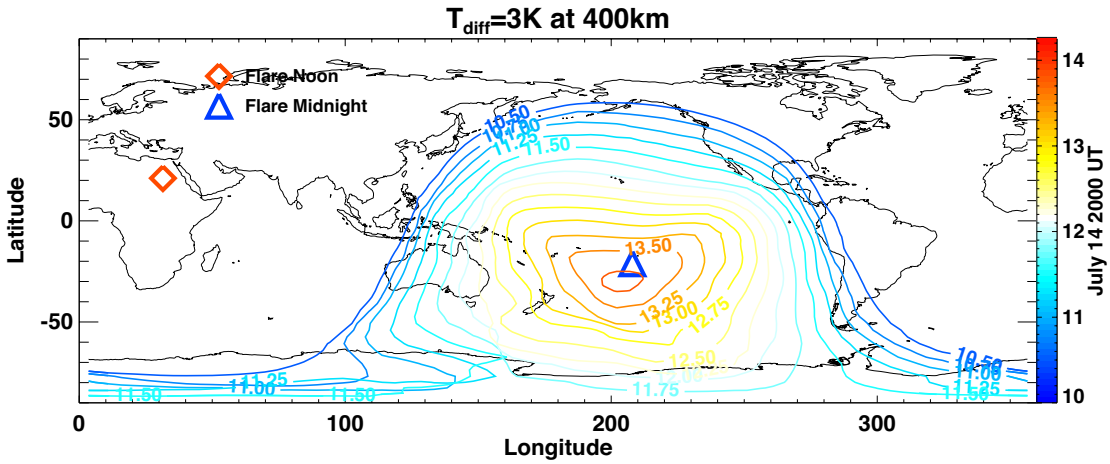

Figure 4.3: Isolines of the different temperature of 5K at 400-km altitude from 1000 UT to 1400 UT on July 14, 2000. The colors of the isolines change with the time. The dark diamond and dark triangle represent the subsolar point and anti-subsolar point at the flare onset (i.e., 1003 UT). The light triangle marks the anti-subsolar point at 1400 UT.

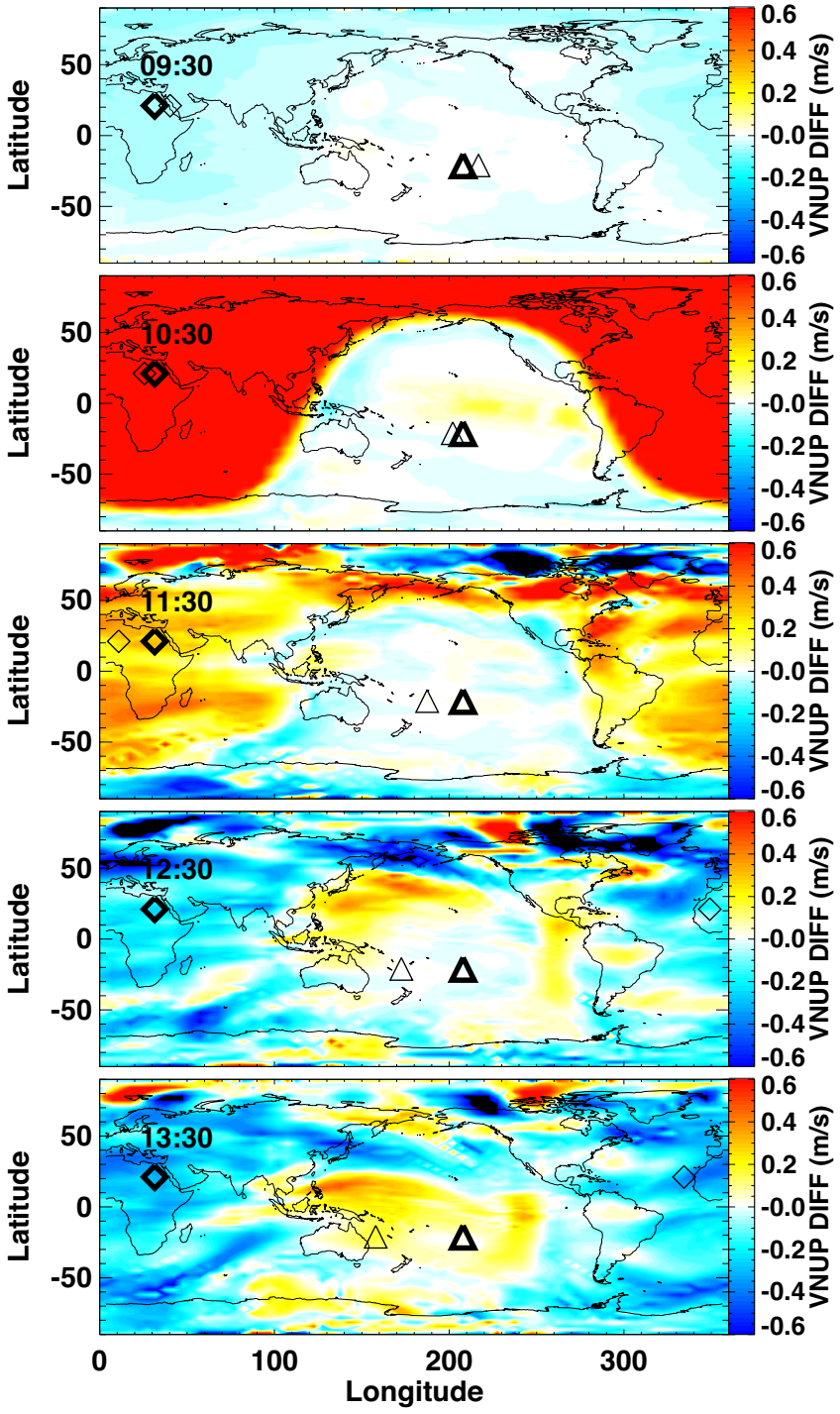

Figure 4.4: The global map of the differenced low-frequency vertical neutral wind at 0930 UT, 1030 UT, 1130 UT and 1330 UT. The thick diamond and the thick triangle mark the subsolar and anti-subsolar points at the time of the plot; the thin triangles mark the anti-subsolar point when the flare occurred (1003 UT).

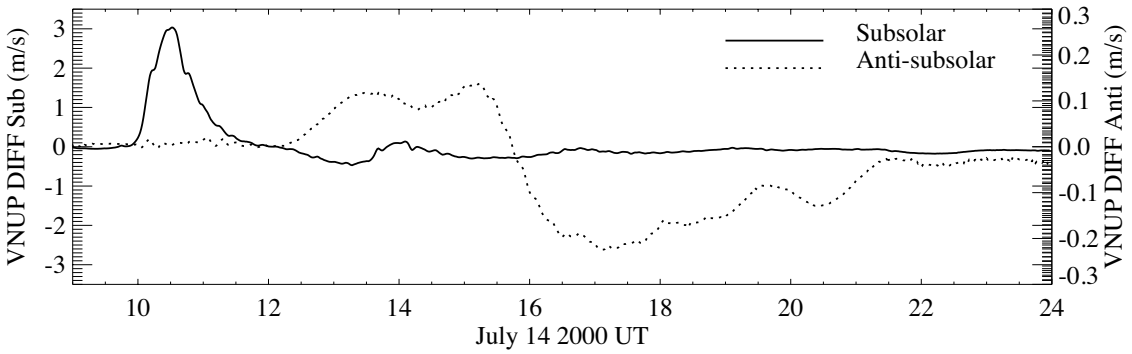

Figure 4.5: The low-frequency component of the vertical winds at flare-time subsolar (solid line) and anti-subsolar (dotted line) points from 0900 UT to 2400 UT on July 14, 2000.

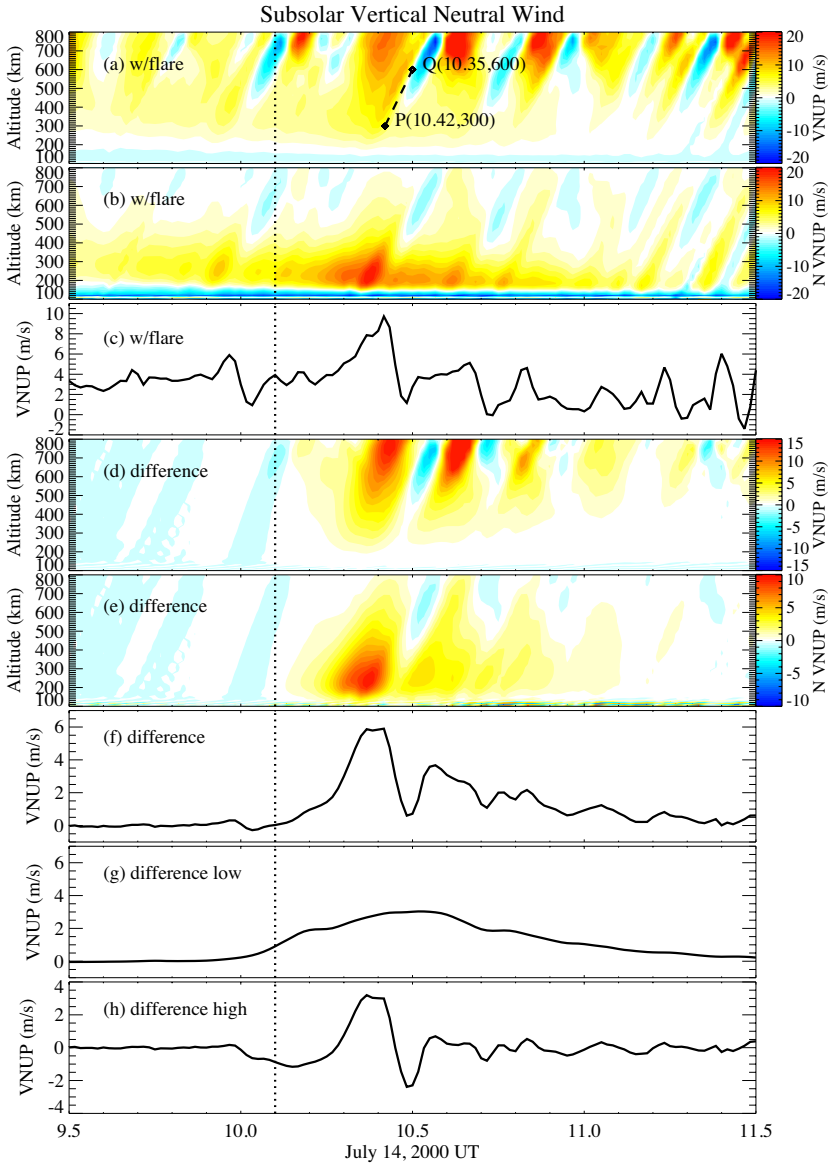

Figure 4.6: The temporal variation of the vertical neutral wind at subsolar point from 0930 UT to 1130 UT on July 14, 2000. The dashed vertical lines at 1003 UT mark the time of the flare onset. (a) the altitudinal profiles of the raw vertical neutral wind from 100 km to 800 km. (b) the altitudinal profiles of the normalized raw vertical neutral wind. (c) the raw vertical neutral wind at 400-km altitude. (d) the altitudinal profiles of the differenced vertical wind. (e) the altitudinal profiles of the normalized differenced vertical wind. (f) the differenced vertical wind at 400-km altitude. (g) the low-frequency component of the differenced vertical wind at 400-km altitude. (h) the high-frequency component of the differenced vertical wind at 400-km altitude.

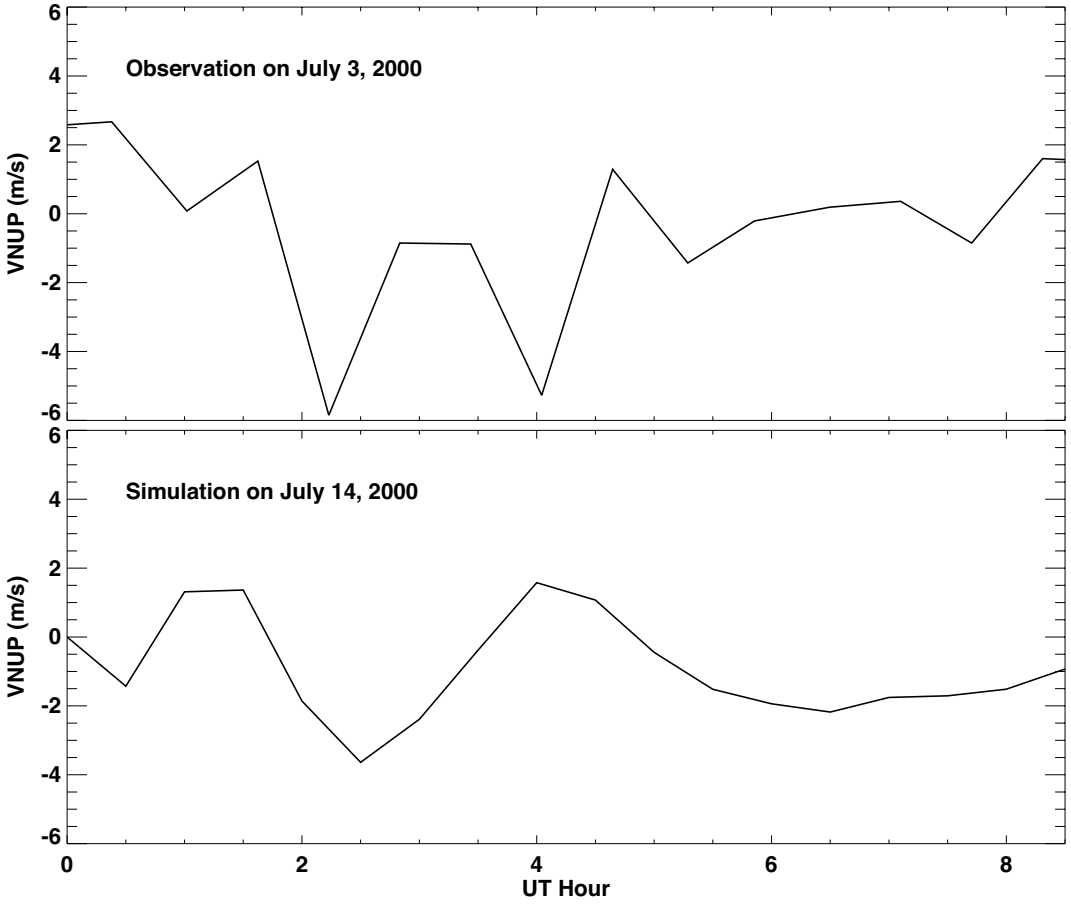

Figure 4.7: The vertical winds observed by the FPI experiment at Arecibo on July 3, 2000 (top) and that simulated in GITM at 300 km altitude at Arecibo on July 14, 2000 (bottom).

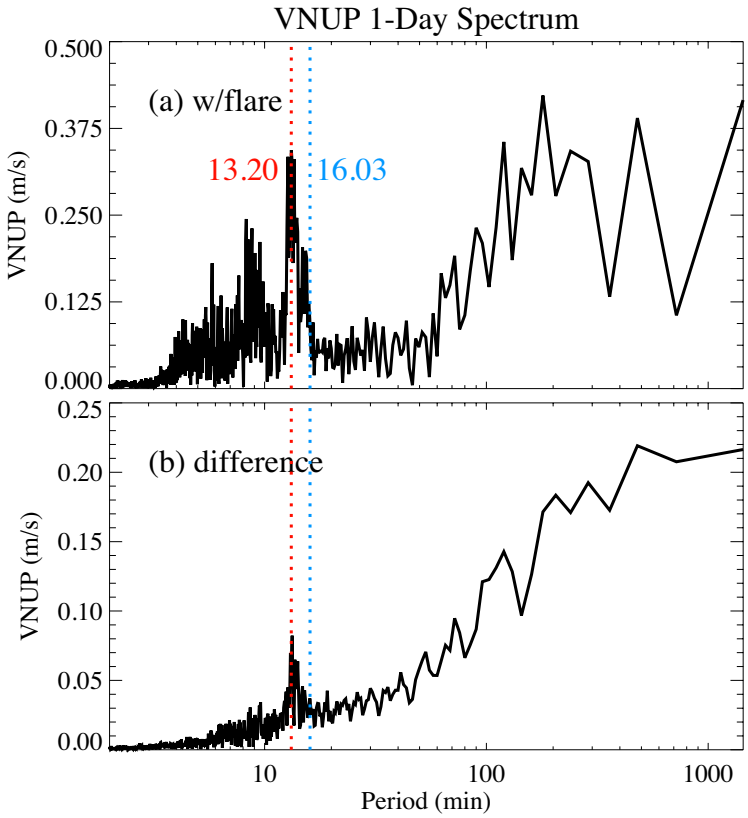

Figure 4.8: The Fast Fourier Transforms (FFT) of the vertical neutral wind at 400-km altitude for one-day period on July 14, 2000. The blue dashed vertical line marks the average buoyancy period of about 16 minutes; the red dashed line marks the period where the vertical wind power lies. The top panel shows the one-day spectrum of the raw vertical neutral wind; the bottom panel shows the differenced vertical neutral wind.

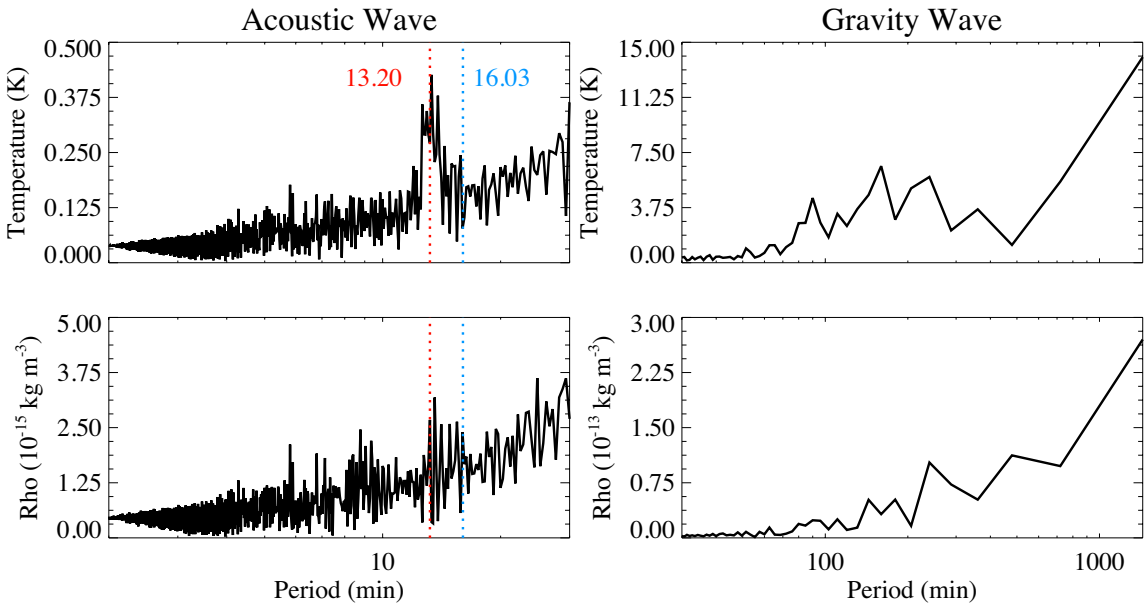

Figure 4.9: The Fast Fourier Transforms (FFT) of the neutral temperature (top row) and density (bottom row) at 400-km altitude for one-day period on July 14, 2000. The left panels show the spectra of the acoustic wave from 0 to 30 minutes, while the right panels show the spectra of the gravity waves from 30 minutes to 24 hours. The blue dashed vertical lines mark the average buoyancy period; the red dashed line marks the period where the vertical wind power lies.

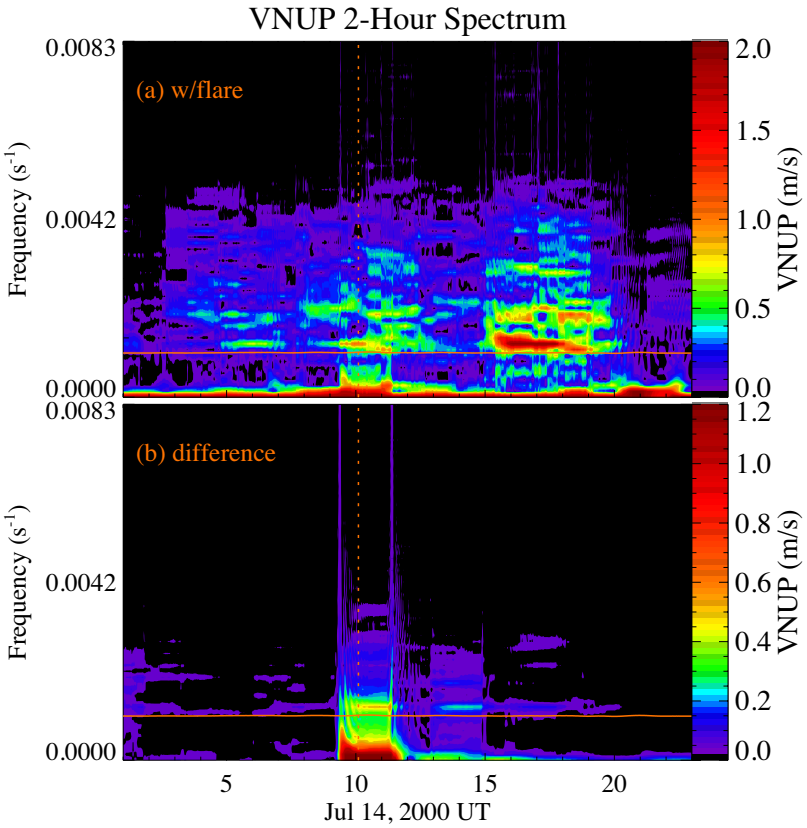

Figure 4.10: Spectrum of the vertical neutral wind at 400-km altitude within every two-hour window. The top panel shows the spectrum of the flare vertical neutral wind; the bottom shows the spectrum of the differenced vertical neutral wind. The red solid line indicates the buoyancy frequency at the centered time of each window. The red dashed line marks the time of the flare onset.

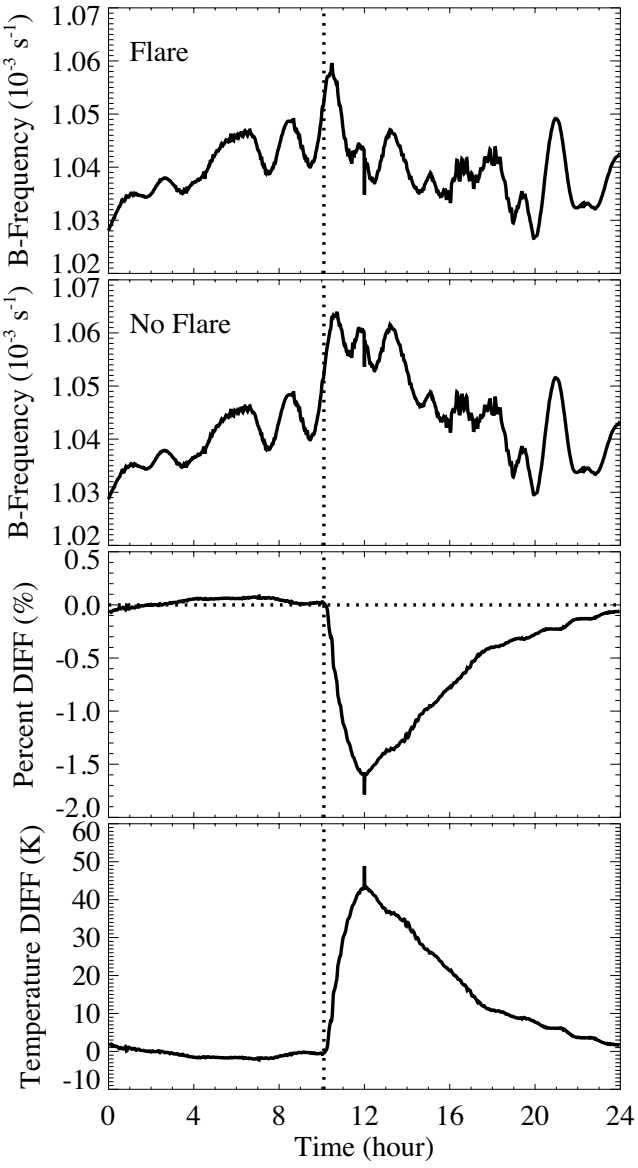

Figure 4.11: Temporal variation of the subsolar buoyancy frequency at 400 km for the flare simulation (top panel) and for the non-flare simulation (second panel) for July 14, 2000. The third and bottom panels show the percent difference in buoyancy frequency and the temperature difference due to the flare respectively. The dashed vertical lines mark the flare onset.

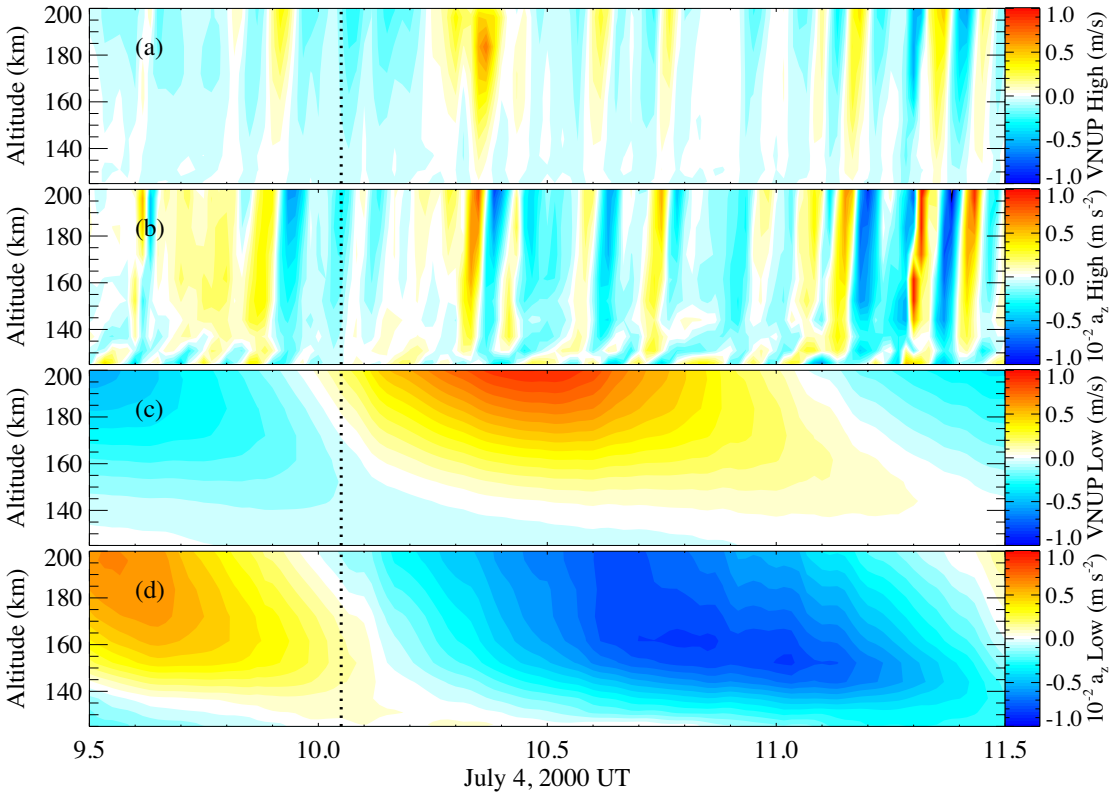

Figure 4.12: Temporal variation of vertical neutral wind zoomed between 125 km and 200 km from 0930 UT to 1130 UT. Panels (a) and (b) show the high-frequency components of the vertical neutral wind and the neutral parcel acceleration respectively. Panels (c) and (d) show the low-frequency components of the vertical neutral wind and the neural parcel acceleration respectively. The dotted lines mark the flare onset.

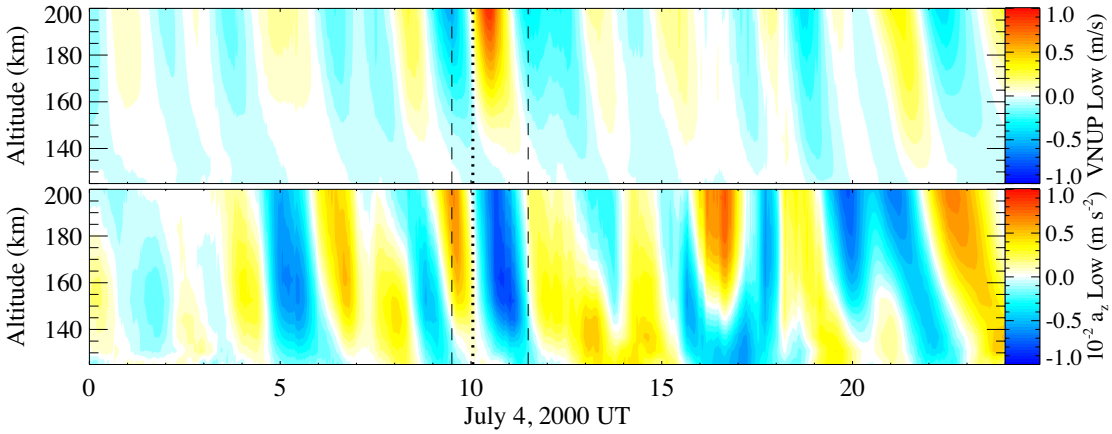

Figure 4.13: The temporal variation of vertical neutral wind between 100 km and 200 km from 0000 UT to 1130 UT. The top and bottom panels show the low-frequency vertical wind and the parcel acceleration respectively. The time frame between the two dashed lines is from 0930 UT to 1130 UT. The dotted lines mark the flare onset.

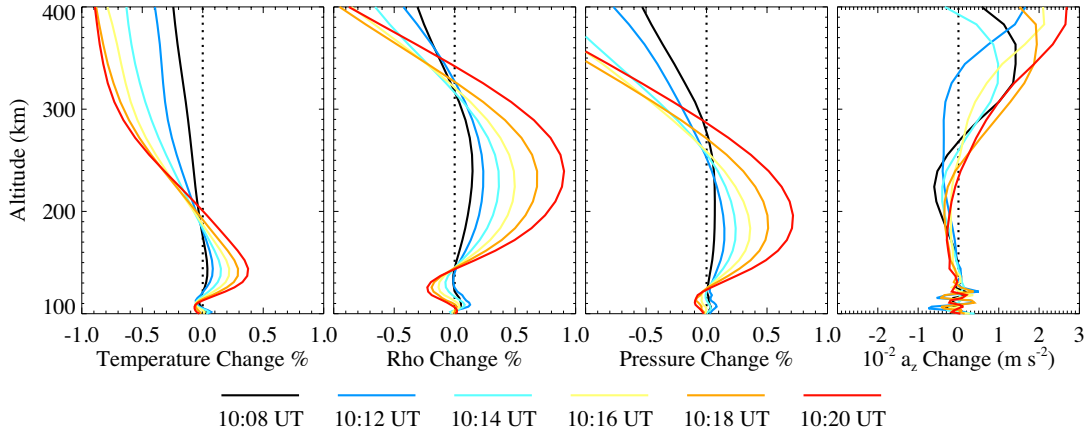

Figure 4.14: From left to right: the vertical profiles of the percentage changes in the subsolar neutral temperature, mass density and pressure, as well as the difference in the vertical acceleration respectively with respect to 10:06 UT. For example, the yellow line in the left most plot shows $\frac{T_{1016}-T_{1006}}{T_{1006}} \times 100\%$, while in the right most plot shows $a_{z1016} - a_{z1006}$. The six colored lines represent six UTs as indicated below. The vertical dashed line marks "zero" change.

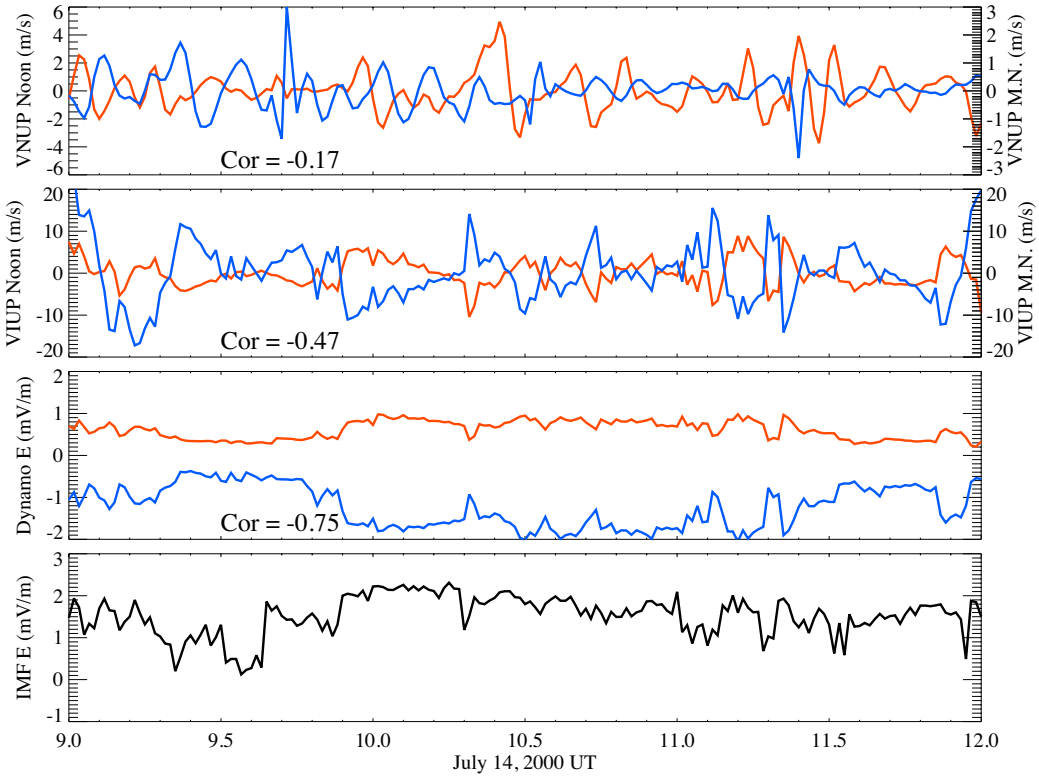

Figure 4.15: The temporal variation of the high-frequency component of the raw vertical neutral wind (top panel), the high-frequency component of the raw vertical ion velocity (second panel) and the zonal dynamo electric field (third panel). The red and blue lines show the data on the subsolar and anti-subsolar sectors respectively. The bottom panel shows the temporal variation of the high-latitude electric field driven by the interplanetary magnetic field.

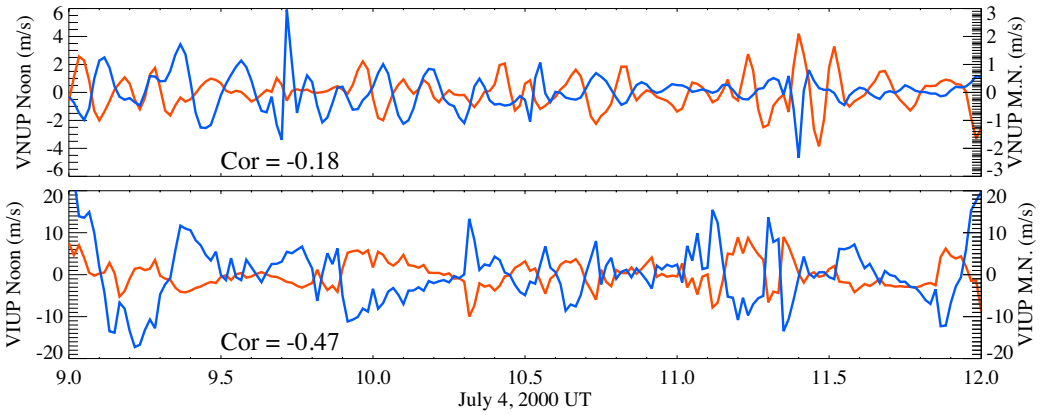

Figure 4.16: The temporal variation of the high-frequency component of the raw vertical neutral wind without the flare (top panel) and the high-frequency component of the raw vertical ion velocity without the flare (bottom panel). The red and blue lines show the data on the subsolar and anti-subsolar sectors respectively.

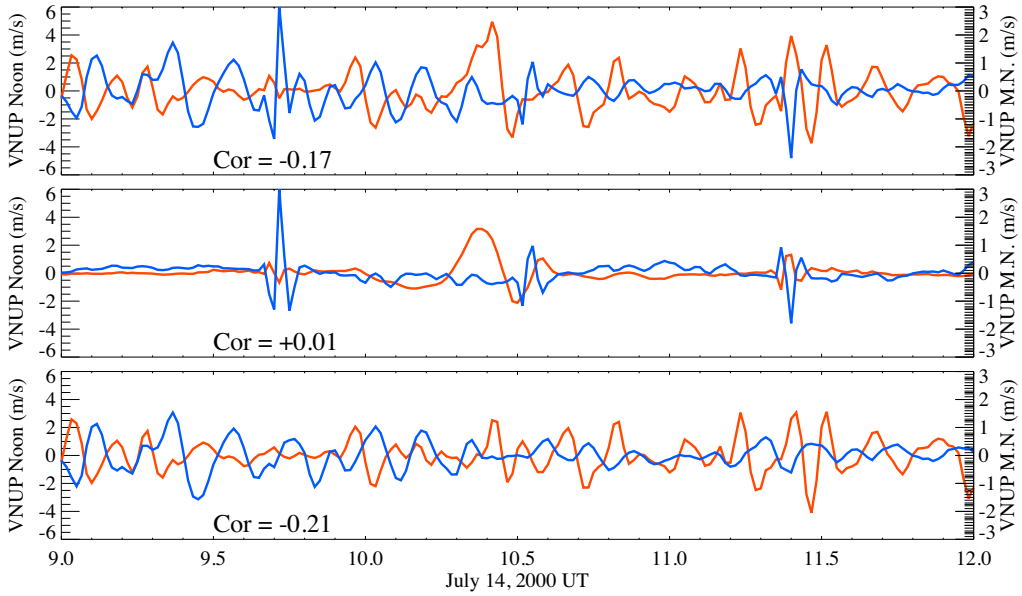

Figure 4.17: The temporal variation of the high-frequency component of the raw vertical neutral wind with dynamo on (top panel), the high-frequency component of the raw vertical neutral wind with dynamo off (second panel), and the difference in high-frequency component of the raw vertical neutral wind between dynamo-on and dynamo-off. The red and blue lines show the data on the subsolar and anti-subsolar sectors respectively.

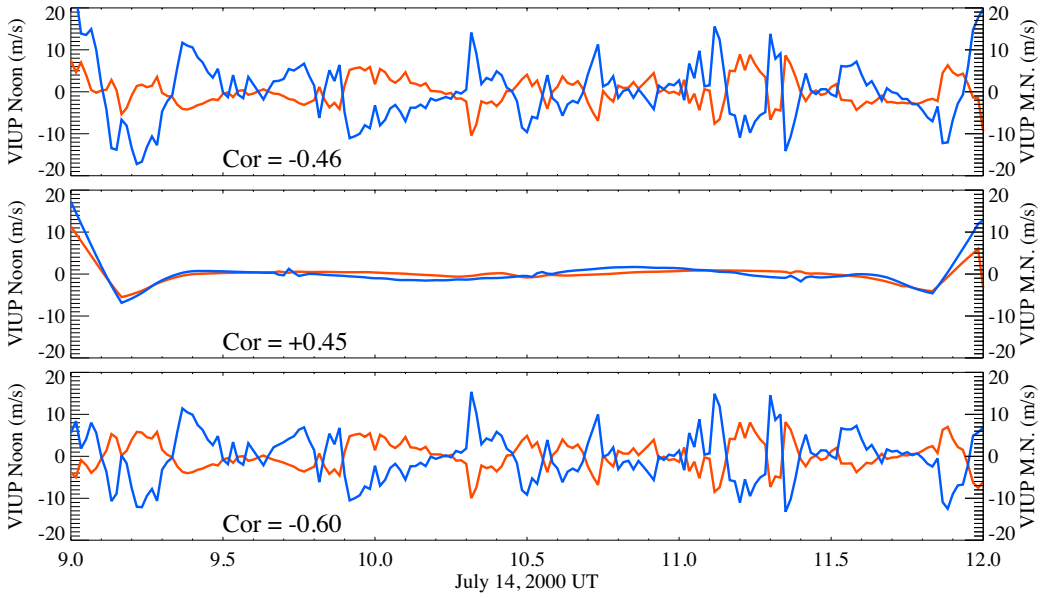

Figure 4.18: The temporal variation of the high-frequency component of the raw vertical ion velocity with dynamo on (top panel), the high-frequency component of the raw vertical ion velocity with dynamo off (second panel), and the difference in high-frequency component of the raw vertical ion velocity between dynamo-on and dynamo-off. The red and blue lines show the data on the subsolar and anti-subsolar sectors respectively.

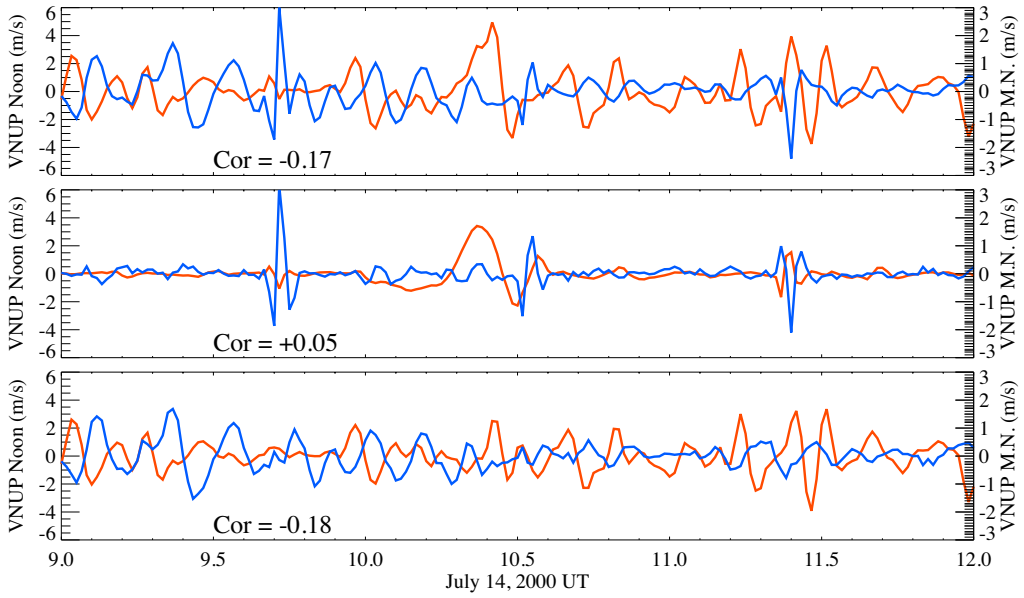

Figure 4.19: The temporal variation of the high-frequency component of the raw vertical neutral wind with the satellite IMF input (top panel), the high-frequency component of the raw vertical neutral wind with the constant IMF input (second panel), and the difference in high-frequency component of the raw vertical neutral wind between the satellite IMF and constant IMF. The red and blue lines show the data on the subsolar and anti-subsolar sectors respectively.

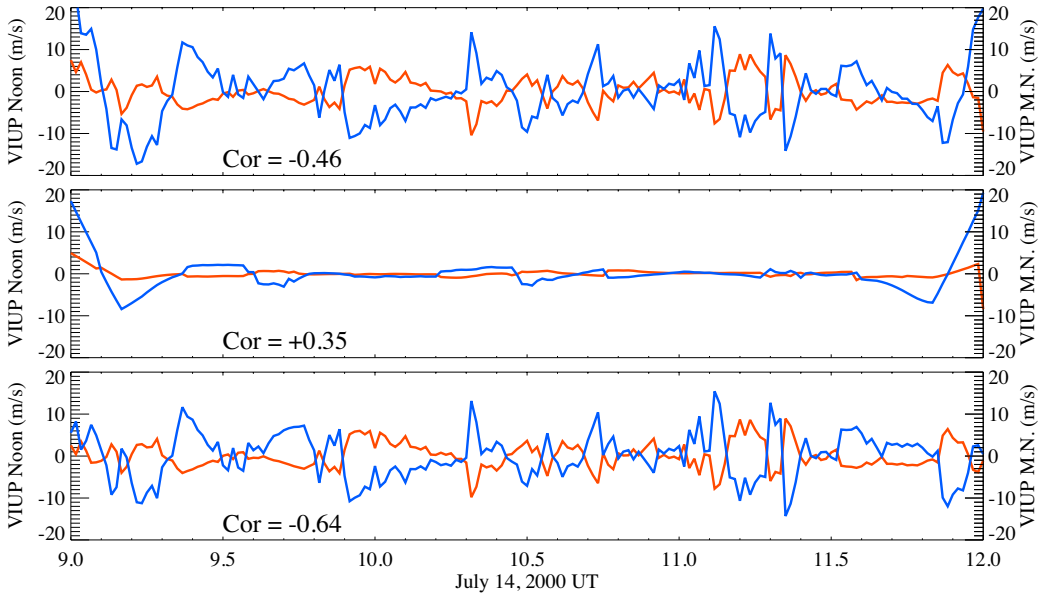

Figure 4.20: The temporal variation of the high-frequency component of the raw vertical ion velocity with the satellite IMF input (top panel), the high-frequency component of the raw vertical ion velocity with the constant IMF input (second panel), and the difference in high-frequency component of the raw vertical ion velocity between the satellite IMF and constant IMF. The red and blue lines show the data on the subsolar and anti-subsolar sectors respectively.

CHAPTER V

Simulating electron and ion temperature in a global ionosphere thermosphere model: validation and modeling an idealized substorm

Electron and ion temperatures control many chemical and physical processes in the ionosphere-thermosphere system. Recently, improved electron and ion energy equations were implemented in GITM. The source energy of the electron temperature (T_e) include thermal conduction, heating due to photoionization, elastic collisions with ions, elastic and inelastic collisions with neutrals, auroral precipitation, and heat flux from inner magnetosphere. The source terms in the ion temperature (T_i) equation include thermal conduction, and elastic collisions with electrons and neutrals. The new implementation of T_e improved the ionospheric density at middle and high latitudes with respect to IRI. The improved GITM also reproduced the diurnal variation in T_e and T_i observed by incoherent scatter radars at low and middle latitudes. The model was used to investigate an idealized substorm statistically described by *Clausen et al.* (2014). It was found that the responses of the E-region N_e and T_e were highly correlated with the variation in auroral hemispheric power. The change of the F-region T_e was correlated with the E-region T_e and N_e, which was consistent with observations. The response of the F-region N_e to the particle precipitation

was delayed by about 30 minutes, and lasted significantly longer than the enhanced precipitation. The variations of T_i in both the E- and F-regions were dominated by IMF-driven ion drifts through frictional energy coupling with the neutrals. It was also found that the increase of the mid-latitude heat flux by one order of magnitude enhanced Te, electron density and TEC by up to 120%, 80% and 80% respectively between dusk and midnight.

5.1 Introduction

Electron and ion temperatures strongly affect the ionosphere-thermosphere system because they control many chemical and physical processes (*Schunk and Nagy*, 2009); therefore, investigating the temporal and spatial variations of the temperatures is important in understanding the coupled system. It has been widely observed that variations exist in the F-region T_e at low and middle latitudes, which are characterized by a "morning overshoot" (*McClure*, 1971; *Oyama et al.*, 1996) and an evening enhancement (*Fukao et al.*, 1991). These morning and evening enhancements are mainly caused by the photoelectron heating of the low-density thermal electrons at dawn and dusk (*Dalgarno et al.*, 1963; *Da Rosa*, 1966b; *Otsuka et al.*, 1998). The morning peak tends to occur in equinox and winter at high solar activity at Millstone Hill, while the evening enhancement is weaker or absent. At Arecibo, the morning enhancement occurs during all seasons, while the evening enhancement only tends to occur in the equinox and winter seasons (*Lei et al.*, 2007). *Truhlik et al.* (2009) found an anti-correlation between the daytime T_e and solar activity at mid-latitudes below 700 km during the equinox and winter, while the nighttime T_e increased with solar activity at low and mid-latitudes at all altitudes. T_e can also be significantly enhanced by heat flux from the plasmasphere in regions with low electron density, especially inside the mid-latitude electron density trough (*Wang et al.*, 2006). The high-latitude T_e is strongly affected by the polar heat flux and field-aligned current (FAC) (*Schunk*

et al., 1986a). A downward polar heat flux can affect the electron density and temperature near the F-region and above (*David et al.*, 2011). Upward (Downward) FACs increase (decrease) T_e at altitudes higher than the F-region maximum (*Zhang et al.*, 2003).

Banks (1967) suggested a transition behavior of T_i, which increased from the neutral temperature near 250 km towards T_e at higher altitudes. The daytime high-latitude T_i generally tends to be higher in summer than in winter, at solar maximum than at solar minimum, and for active magnetic conditions than for quiet conditions (*Schunk and Sojka*, 1982). They also found that an upward heat flow from the lower ionosphere due to the meridional electric fields acts to raise T_i at high altitudes. The high-latitude F-region T_i tends to be enhanced by ion-neutral friction heating during geomagnetic storms, leading to an increase in T_e due to the ion-electron energy transfer (*Wang et al.*, 2006). T_i is strongly related to the ion-neutral collision frequencies that affect the heat transfer rates between ions and neutrals. The ion energy equation has been approximated by a balance between the energy exchange and frictional heating with neutrals below 400 km at high latitudes (*Thayer and Semeter*, 2004).

Efforts have been made in modeling T_e and T_i in the ionosphere since the 1980s. The Time-Dependent Ionosphere Model (TDIM) solves the electron energy equation considering thermal conduction, thermoelectric transport, ion Joule heating, heating due to photoelectrons and auroral electrons, and collisions with the thermal ions and neutrals (*Schunk et al.*, 1986a; *Schunk*, 1988; *Sojka*, 1989; *David et al.*, 2011). The Coupled Thermosphere-Ionosphere-Plasmasphere (CTIP) model solves the time-dependent T_e and T_i for O^+ and H^+ in field-aligned coordinates. The sources terms include heating due to collisional interactions with other species, adiabatic heating/cooling and thermal conductivity (*Millward et al.*, 1996). SAMI2 solves similar ion and electron energy equations in field-aligned coordinates with a semi-implicit method (*Huba et al.*, 2000). T_e in the Thermosphere-Ionosphere Nested Grid (TING)

model and the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIE-GCM) is solved assuming a steady state with vertical heat conduction and heating due to photoelectrons and particle precipitation, as well as cooling to ions and neutrals (*Maute*, 2011; *Wang et al.*, 2006).

Since the early 1980s, radar and rocket observations have been used to study the ionospheric dynamics during substorms (*Baumjohann et al.*, 1981; *Inhester et al.*, 1981; *Opgenoorth et al.*, 1990). *Maehlum et al.* (1984) found that the F-region electron temperature was correlated with the E-region density based on the EISCAT observations during an intense aurora. They suggested that the F-region could be heated by current driven instabilities associated with intense particle precipitation. Rocket observation showed that the ionospheric electron temperature was significantly higher than would be due to collisional heating with precipitating particles and it was suggested that some wave-particle interaction contributed to the extra heating (*Svenes et al.*, 1992b). The F-region T_i measured by EISCAT UHF-radar was found to be dominated by the local electric field (*Svenes et al.*, 1992a). During the expansion phase of an isolated substorm, both T_e and N_e were strongly enhanced (*Yeoman et al.*, 2000). Although the thermal response of the ionosphere during a substorm has been greatly studied, there have been few studies using global ionosphere thermosphere models to investigate the response of the ionospheric temperatures during substorms. In this study, GITM with the recently implemented ionospheric temperature model, was used to explore the response of the ionosphere to an idealized substorm (*Clausen et al.*, 2014; *Liu and Ridley*, 2015) and the effect of the topside electron heat flux.

5.2 Model Description

In this study, the *Weimer* (2005) model was used for the high-latitude electric fields, and the *Fuller-Rowell and Evans* (1987) model was employed for produce the auroral precipitation patterns.

5.2.1 Electron Temperature

Instead of assuming a steady state electron temperature, as was done in earlier versions of GITM, the new model solves for a time-dependent electron energy equation. This allows a (possibly) more precise description of the temporal variations of T_e. If chemical reactions and viscous heating of the electron gas are neglected, the electron energy equation can be expressed as (*Schunk and Walker*, 1970):

$$\frac{3}{2}\kappa n_e \frac{\partial T_e}{\partial t} = -\kappa n_e T_e \bigtriangledown \cdot \boldsymbol{u_e} - \frac{3}{2} N_e \kappa \boldsymbol{u_e} \cdot \bigtriangledown T_e - \bigtriangledown \cdot \boldsymbol{q_e} + \Sigma Q_e, \tag{5.1}$$

where κ is the Boltzmann constant, n_e is the electron density, T_e is the electron temperature, $\boldsymbol{u_e}$ is the electron velocity, $\boldsymbol{q_e}$ is the heat flow vector, t is time and z is altitude. The terms on the right-hand side of Equation (5.1) represent (from left to right) the adiabatic expansion, heat advection, and the divergence of the electron heat flow vector respectively. ΣQ_e is the sum of all the local heating and cooling rates, respectively. In this model, the adiabatic heating/cooling is neglected. The heat flow vector is expressed as

$$\boldsymbol{q_e} = -\chi \boldsymbol{J}_{\parallel} - \lambda_e \bigtriangledown T_e. \tag{5.2}$$

where χ is the thermoelectric coefficient, $\chi = \frac{5}{2}\kappa T_e/e$ (*Schunk and Walker*, 1970) and λ_e is the electron thermal conductivity (given in the Appendix). This model also included the thermoelectric heating along the field lines above 45deg Magnetic Latitude. The field aligned current is based on the divergence of the ion and electron flows in the plane perpendicular to the field lines. Considering thermal advection only in the radial direction, the electron energy equation implemented in the new model can be simplified to:

$$\frac{3}{2}\kappa n_e \frac{\partial T_e}{\partial t} = -\frac{3}{2}\kappa n_e u_e \frac{\partial T_e}{\partial r} + Q_{J_{//}} + Q_{conduction} + \Sigma Q_e. \tag{5.3}$$

where r is the unit vector in the radial direction in the Earth's spherical coordinates. The terms on the right side represent (from left to right) the thermal advection in the radial direction, thermoelectric heating, thermal conduction and the sum of other local heating/cooling sources. The field aligned thermal conduction can be written as:

$$Q_{conduction} = \frac{\partial}{\partial s}(\lambda_e \frac{\partial T_e}{\partial s}) = \lambda_e \frac{\partial^2 T_e}{\partial s^2} + \frac{\partial \lambda_e}{\partial s}\frac{\partial T_e}{\partial s} \tag{5.4}$$

Since the electron thermal conduction primarily occurs along magnetic field lines (*Schunk et al.*, 1986a), there has been no global ionosphere thermosphere model that includes the thermal conduction perpendicular to field lines, to the authors' knowledge. In this model, thermal conduction both along and perpendicular to the field lines are implemented. Projecting the field-aligned conduction to spherical coordinates gives:

$$\begin{aligned} Q_{conduction} = \sin^2\alpha(\lambda_e \frac{\partial^2 T_e}{\partial r^2} + \frac{\partial T_e}{\partial r}\frac{\partial \lambda_e}{\partial r}) \\ +\lambda_e \cos^2\alpha(\frac{1}{r^2}\cos^2\beta\frac{\partial^2 T_e}{\partial \theta^2} + \frac{\sin^2\beta}{r^2\cos^2\theta}\frac{\partial^2 T_e}{\partial \phi^2} + \frac{2\sin\beta\cos\beta}{r^2\cos\theta}\frac{\partial^2 T_e}{\partial\theta\partial\phi} - \frac{\sin\beta\cos\beta\sin\theta}{r^2\cos^2\theta}\frac{\partial T_e}{\partial\phi}) \\ +\cos^2\alpha(\cos^2\beta\frac{1}{r^2}\frac{\partial\lambda_e}{\partial\theta}\frac{\partial T_e}{\partial\theta} + \sin^2\beta\frac{1}{r^2\cos^2\theta}\frac{\partial\lambda_e}{\partial\phi}\frac{\partial T_e}{\partial\phi} + \sin\beta\cos\beta\frac{1}{r^2\cos\theta}(\frac{\partial\lambda_e}{\partial\theta}\frac{\partial T_e}{\partial\phi} + \frac{\partial\lambda_e}{\partial\phi}\frac{\partial T_e}{\partial\theta})) \\ +\sin\alpha\cos\alpha\cos\beta\frac{1}{r}(\frac{\partial T_e}{\partial\theta}\frac{\partial\lambda_e}{\partial r} + \frac{\partial T_e}{\partial r}\frac{\partial\lambda_e}{\partial\theta}) + \sin\alpha\cos\alpha\sin\beta\frac{1}{r\cos\theta}(\frac{\partial T_e}{\partial\phi}\frac{\partial\lambda_e}{\partial r} + \frac{\partial T_e}{\partial r}\frac{\partial\lambda_e}{\partial\phi}), \end{aligned} \tag{5.5}$$

where α and β are the magnetic dip and declination angle, respectively. θ and ϕ are the latitude and longitude in the Earth's spherical coordinates, respectively. The cross terms between the radial direction and the horizontal plane, i.e., the last two terms in Equation (5.5), were ignored for simplicity.

Considering the field-aligned currents are prominent in high latitudes, the heat-

ing/cooling due to field-aligned current as considered above 45deg Magnetic Latitude. The field aligned current $J_{//}$ was calculated as:

$$J_{//} = \int_{r_0}^{r_1} \bigtriangledown \cdot J_{\perp}, \tag{5.6}$$

where r_0 and r_1 are the lower and upper boundary of the ionosphere, respectively. $J_{\perp}$ is the current perpendicular to the field lines, and $J_{\perp} = n_e|e|(u_{i\perp} - u_{e\perp})$. Assuming the field aligned current is mainly contributed by electron flow and using the relation $J_{//} = -n_e|e|u_e$, the thermal advection term as expressed as a function of $J_{//}$:

$$-\frac{3}{2}\kappa n_e u_e \frac{\partial T_e}{\partial r} = \frac{3}{2}\kappa \frac{J_{//}}{|e|}\frac{\partial T_e}{\partial r}, \tag{5.7}$$

and was included above 45deg Magnetic Latitude.

The local heating/cooling rates are given by:

$$\Sigma Q_e = Q_{phe} + Q_{auro} + Q_{ei} + [Q_{en}]_{elas} + [Q_{en}]_{inelas}, \tag{5.8}$$

where Q_{phe} is heating due to photoelectrons, Q_{auro} is heating due to auroral ionization, Q_{ei} is the heat transfer between electrons and ions due to elastic collisions, and $[Q_{en}]_{elas}$ and $[Q_{en}]_{inelas}$ represent cooling rates due elastic and inelastic collisions with neutrals. The new model adds cooling due to elastic collisions with neutrals and heating due to auroral ionization, and calculates the photoelectron heating using an improved parameterization model of photoelectron heating efficiency described by *Smithtro and Solomon* (2008), which allows for a change in the heating efficiency as a function of altitude due to the changing of the ratio between the electron and neutral densities. The photoelectron heating can be expressed as:

$$Q_{phe} = \epsilon P_{photo}, \tag{5.9}$$

where ϵ is a thermal electron heating efficiency and P_{photo} is the local photoionization rate (*Swartz and Nisbet*, 1972). The heating efficiency is specified by an improved parameterization derived by *Smithtro and Solomon* (2008):

$$\begin{aligned}\epsilon = exp(5.342 + 1.056x - 4.392 \times 10^{-2}x^2 - 5.900 \times 10^{-2}x^3 \\ -9.346 \times 10^{-3}x^4 - 5.755 \times 10^{-4}x^5 - 1.249 \times 10^{-5}x^6),\end{aligned} \tag{5.10}$$

where

$$x = ln(R), \quad R = \frac{n_e}{n(N_2) + n(O_2) + n(O)}. \tag{5.11}$$

This parameterization performs over a wide altitude range of ∼100 km to ∼ 750 km (*Solomon et al.*, 1988; *Solomon and Abreu*, 1989). The domain of GITM, from ∼90 km to ∼700 km, is basically within its applicable range. The maximum of the ratio R in Equation (5.11) is limited to 10 in the model, which rarely, if ever, actually happens, even at high solar activity. The auroral heating rate is expressed as

$$Q_{auro} = \epsilon P_{auro}, \tag{5.12}$$

where P_{auro} is the auroral ionization rate, and the heating efficiency ϵ is the same as shown in Equation (5.10).

The heat transfer due to elastic collisions with ions is given by:

$$Q_{ei} = 3\kappa n_e m_e \sum_t \nu_{et} \frac{(T_i - T_e)}{m_e + m_t}, \tag{5.13}$$

where m_e is the electron mass, and T_i is the ion temperature (note that GITM only has a single bulk ion temperature). The Coulomb collision frequency (ν_{et}) between electrons and ion species t is given in Appendix A.

The heating term due to collisions with neutrals consists of two parts: elastic

collisions and inelastic collisions. The elastic collision term is given by:

$$[Q_{en}]_{elas} = 3\kappa n_e m_e \sum_s \nu_{es} \frac{T_n - T_e}{m_e + m_s}, \tag{5.14}$$

where T_n is the neutral temperature and ν_{es} is the collision frequency between electrons and neutral species s and is given in Appendix A. This elastic collision term is almost always a cooling term, due to the fact that the neutrals are almost always cooler than the electrons. The corresponding heating rate for neutrals due to elastic collisions with electrons is a source term in the neutral energy equation.

The inelastic collision term is given by:

$$[Q_{en}]_{inelas} = L_e(N_2)_{rot} + L_e(O_2)_{rot} + L_e(N_2)_{vib} + L_e(O_2)_{vib} + L_e(O)_{fine} + L_e(O(^1D)), \tag{5.15}$$

where $L_e(N_2)_{rot}$ and $L_e(O_2)_{rot}$ are the cooling rates due to the rotational excitation of the molecular neutrals N_2 and O_2, $L_e(N_2)_{vib}$ and $L_e(O_2)_{vib}$ are due to vibrational excitation of N_2 and O_2, $L_e(O)_{fine}$ and $L_e(O(^1D))$ are due to the fine structure and the excitation of atomic oxygen to the $O(^1D)$ state (given in Appendix A).

5.2.2 Ion Temperature

The old T_i calculation in GITM was based on a statistical formula assuming a balance between the neutral energy exchange term and the neutral frictional heating term (*Thayer and Semeter*, 2004). In the new model, the physical ion energy equation including thermal diffusion and collisional heat transfer between the ions and the electrons and neutrals is implemented. All the ion species sharing the same grid volume are assumed to have the same temperature. The ion temperature equation can be expressed as:

$$\frac{3}{2}\kappa n_i \frac{\partial T_i}{\partial t} = -\kappa n_i T_i \bigtriangledown \cdot \boldsymbol{u_i} - \frac{3}{2} N_i \kappa \boldsymbol{u_i} \cdot \bigtriangledown T_i - \bigtriangledown \cdot \boldsymbol{q_i} + \Sigma Q_i. \tag{5.16}$$

In GITM, the adiabatic expansion and heat advection were neglected. Also, considering field-aligned currents were mainly driven by electron motion, the thermoelectric heating was ignored in GITM. Thus, the heat flow vector $\boldsymbol{q_i}$ is:

$$\boldsymbol{q_i} = -\lambda_i \bigtriangledown T_i, \tag{5.17}$$

where λ_i is the ion thermal conductivity given in Appendix B.

Similar to electrons, the ion thermal conduction was considered along the magnetic field, both in the vertical and horizontal directions. With these assumptions, the ion energy equation can be written as:

$$\frac{3}{2}\kappa n_i \frac{\partial T_i}{\partial t} = Q_{conduction} + \Sigma Q_i. \tag{5.18}$$

where n_i is the total ion density (equal to the electron density) and T_i is the ion temperature. The ion thermal conduction is in the same format as the electron thermal conduction in Equation (5.5). Q_i represents the total local ion heating/cooling rate, which is the sum of the heat transfer due to collisions with electrons and with neutrals.

The elastic collisional heating of the ions by the electrons is:

$$Q_{ie} = 3\kappa \sum_t \frac{n_t m_t \nu_{te}}{m_t + m_e}(T_e - T_i), \tag{5.19}$$

where the subscript t refers to the ion species. Given that the collision frequencies satisfy the relationship (*Schunk and Nagy*, 2009):

$$n_t m_t \nu_{te} = n_e m_e \nu_{et}, \tag{5.20}$$

Q_{ie} can be further expressed as

$$Q_{ie} = 3n_e m_e \kappa (T_e - T_i) \sum_t \frac{\nu_{et}}{m_e + m_t}. \tag{5.21}$$

The heating term accounting for collisions with neutrals consists of two parts: collisional heat transfer and frictional heating:

$$Q_{in} = \sum_t n_t m_t \sum_k \frac{\nu_{tk}[3\kappa(T_n - T_i) + m_k(u_n - u_i)^2]}{m_t + m_k} \tag{5.22}$$

where n_t and m_t are the number density and mass of ion species t, u_n and u_i are the neutral and ion velocities, and the subscripts t and k denotes the ion and neutral species respectively. The ion-neutral collision frequencies are given in Appendix B. The corresponding heating rate for the neutrals due to collisions with the ions was added to the heating sources in the neutral energy equation.

5.2.3 Numerical Method

The energy equations are solved using a semi-implicit method, which was described by *Huba et al.* (2000). Schematically, the model uses the backward difference at the current time t^n and second-order central difference for the spatial derivative at altitude z_j in the conduction calculation at t^n. Each source term with a linear dependence on the temperature is split into two parts, one with a linear dependence on the variable T_j^n being solved, and the other without the linear dependence. The linear terms are evaluated at t^n, while the other terms are evaluated at the previous time step t^{n-1}. This semi-implicit method performs well in limiting the stiff source terms, and is numerical stable and convergent.

Both T_e and T_i are set equal to the neutral temperature at the bottom boundary. T_e has a heat flux boundary at the top, such that:

$$\frac{T_j^n - T_{j-1}^n}{\delta z} = \frac{F_e}{\lambda_e}, \tag{5.23}$$

where F_e is the heat flux rate. The upper boundary of T_i is set as a constant gradient boundary condition (i.e., only local source terms and no heat flux):

$$(\frac{\partial T^n}{\partial z})_j = (\frac{\partial T^n}{\partial z})_{j-1}. \tag{5.24}$$

Table 5.1 and 5.2 present a comparison of the implementation of T_e and T_i between the old and new model. Four advantages of the new model include: (a) it is capable of simulating T_e stably using the same time step as the thermosphere (1-3s); (b) it incorporates the thermoelectric heating in a global I-T model for the first time; (c) it incorporates the horizontal component of the heat conduction along the field lines for T_e and T_i; (d) it allows energy coupling self-consistently between electrons, ions and neutrals.

Table 5.1: Comparison of T_e Calculation between New and Old Methods

Sources(Method) \ Model	NEW	OLD
Time Derivative	$\surd$	$\times$
Vertical Conduction	$\surd$	$\surd$
Horizontal Conduction	$\surd$	$\times$
Photoelectron Heating	$\surd$	$\surd$
Auroral Ionization Heating	$\surd$	$\times$
Electron-Ion Collisional Heating	$\surd$	$\surd$
Electron-Neutral Elastic Collisional Heating	$\surd$	$\times$
Electron-Neutral Inelastic Collisional Heating	$\surd$	$\surd$
Thermoelectric Heating	$\surd$	$\times$
Magnetospheric Heat Flux	$\surd$	$\surd$
Two-way Coupling with Ions	$\surd$	$\times$
Two-way Coupling with Neutrals	$\surd$	$\times$
Scheme for Vertical Conduction	Implicit	Implicit
Scheme for Other Source Terms	Implicit	Explicit

Table 5.2: Comparison of T_i Calculation between New and Old Methods

Sources(Method) \ Model	NEW	OLD
Time Derivative	√	×
Vertical Conduction	√	×
Horizontal Conduction	√	×
Ion-Electron Collisional Heating	√	×
Ion-Neutral Elastic Collisional Heating	√	√
Two-way Coupling with Electrons	√	×
Two-way Coupling with Neutrals	√	√
Scheme	Implicit	Steady-State

5.3 Validation

5.3.1 Validation with IRI

Figure 5.1 shows comparisons of N_e (top), T_e (middle) and T_i (bottom) using the old model (left), the new model (middle) and the International Reference Ionosphere (IRI) model (*Bilitza et al.*, 2014) (right) at an altitude of 400 km at 00:00 UT on Dec 23, 2012. The simulation was driven by a constant $F_{10.7}$ index of 110.3 and constant solar wind conditions: V_x = 400 km/s and B_x = 0 nT, B_y = 0 nT and $B_z = -2$ nT. The T_e produced by the new model was significantly larger than the previous model, especially in the middle and high latitudes on the dayside. Moreover, due to an increase in the ionospheric scale height and a decrease of recombination rates, the magnitude of the ion density in the new model was also increased and had a better agreement with that in IRI. However, the general distributions of the ion densities were similar between the previous and the improved model. The new model reproduced the diurnal variation in the F-region T_e at low latitudes, which was characterized by a morning and an evening enhancement. The morning peaks were more pronounced in the winter hemisphere than in the summer hemisphere in both the new model and in IRI, which agrees with *Oyama et al.* (1996). Compared to IRI, the morning enhancement of the new T_e was approximately 500 K cooler and had a narrower zonal extension at low latitudes. Both the new GITM and IRI show

a cooler region of T_e in the dayside equatorial region. This feature was caused the inverse correlation between T_e and N_e in the F-region (*Mahajan*, 1977; *Bilitza*, 1991). Specifically, the cooling of the electrons exceeds the photoionization heating when N_e increases to a certain level (*Schunk and Nagy*, 1978a). T_i in the new model was higher than it was in the old model in both the dayside middle and high latitude regions. The ion Joule heating effect was enhanced in the polar regions in the new model. The new T_i has a similar magnitude as IRI on the low-latitude dayside and the south polar region, but was about 500 K cooler in the north polar region. The difference in T_i between the south and north polar regions in GITM was due to the seasonal effect, i.e., the asymmetric solar insolation and geomagnetic forcing (*Rees and Fuller-Rowell*, 1989). There was more N_e in IRI in the northern polar region than in GITM, which would cause a difference in the Joule Heating rate and T_i.

Figure 5.2 shows a comparison of electron density (top), electron temperature (middle) and ion temperature (bottom) between the old GITM (left), the new GITM (middle) and IRI (right) at a longitudinal slice of 180° at 00 UT (i.e., at noon local time). In both the new GITM and IRI, a region filled with denser and cooler electrons existed at low latitudes. However, the topside ionosphere in the new GITM was approximately 500 K cooler than in IRI in this region. This was due to the inverse relation between T_e and N_e: a denser ionosphere existed above 400 km in GITM. Once again, the ion density in the improved model was raised to a more comparable value to IRI due to the increase in T_e. The distribution of T_i in GITM and IRI look very consistent except for the northern polar region, which could be due to the difference in N_e, thus causing a difference in Joule heating and T_i in this region between the two models.

5.3.2 Validation with Incoherent Scatter Radars

A time period from April 13th and 14th, 2013 was simulated with the new model when observations from four incoherent scatter radar (ISR) sites were available. The four ISR sites were Arecibo (18.4°N, 66.6°W), Jicamarca ISR (11.9°S, 76.8°W), Millstone Hill ISR (40.6°N, 71.5°W) and Sondrestrom ISR (67.0°N, 50.7°W). ISRs are capable of measuring the altitudinal profiles of electron density, electron temperature, and ion temperature, to heights of several hundred kilometers by detecting the incoherent backscatter from free electrons in the ionosphere (*Gordon*, 1958b). The simulation was driven by the hemispheric power (HP) obtained by the National Oceanic and Atmospheric Administration (NOAA) POES satellites, as well as the solar wind and interplanetary magnetic field data measured by the Advanced Composition Explorer (ACE), delayed by an appropriate amount of time. The IMF, hemispheric power, and solar wind velocity for April 13 and 14, 2013 are shown in Figure 5.3. The geomagnetic condition was relatively quiet on April 13th, 2013, with hemispheric power less than 10 GW most of the day, solar wind speed about 400 km/s and the magnitude of IMF below 5 nT. On April 14, the hemispheric power and solar wind speed increased to about 30 GW and to about 500 km/s early in the day. Both B_y and B_z turned more positive with larger variability. The $F_{10.7}$ indexes were 125 and 117 on April 13th and 14th, respectively. Figure 5.4 shows the electron heat flux pattern at the top boundary at 12:00 UT. This pattern was fixed in solar magnetic coordinates but moved in geographic coordinates through the day (*Ridley et al.*, 2006). The pattern was constructed based on the modeled energy deposition rates from the ring current into the ionospheric thermal electrons during the June 1991 storm (*Liemohn et al.*, 2000).

Figure 5.5 shows a comparison between GITM, IRI and measurement by the Arecibo ISR from 100 km to 650 km on April 13th and 14th, 2013. The missing data in the observations were filled by an altitudinal linear interpolation. GITM reproduced

the diurnal variation of T_e observed by the ISR, but with a weaker morning peak. In both GITM and the ISR data, the morning enhancement occurred at the same local time at all altitudes. However, the local time of the morning enhancement varied with altitudes in IRI. The T_e in IRI was about 500 K higher than GITM and the ISR data. GITM, IRI and the ISR data had a similar diurnal variation of T_i below 400 km: T_i began to raise around 6 LT, peaked around 16 UT and decreased to the nightside temperature around 20 LT. Above 500 km, GITM underestimated T_i by about 300 K compared to IRI and the ISR data. Also, a distinct morning enhancement in T_i was present in GITM, but not IRI or the ISR data.

A comparison between GITM, IRI and the Jicamarca ISR on April 13, 2013 is shown in Figure 5.6. A morning enhancement in T_e existed between 300 km and 500 km in GITM, ISI and ISR. However, the morning enhancement in IRI was about 500 K cooler than GITM and the ISR data. Also, the morning enhancement occurred approximately at the same local time throughout the altitudes in GITM. In IRI, it occurred earlier at higher altitudes, which was opposite of the ISR observation.There were scattered structures of enhanced T_e on the nightside in the observations, which were not shown in GITM and IRI. This could be equatorial ionospheric perturbations caused by physics that may not be included in the models, or uncertainty in the measurements. The dayside ion temperatures were comparable between the models and the observation. Similar to T_e, structures of high T_i on the nightside were measured by the radar, but were not present in the GITM or IRI.

T_e in the F-region highly depends on the downward electron heat flux from the magnetosphere throughout the mid- and high-latitudes (*Schunk et al.*, 1986a; *Bekerat et al.*, 2007; *David et al.*, 2011). Two simulations with different heat fluxes specified were conducted to compare the model with mid-latitude observations by the Millstone Hill ISR. Figure 5.7 shows the diurnal pattern of the heat flux: the small (left y-axis) and large (right y-axis) share the same pattern but with a one order of magnitude

difference. The large heat flux was closer to a realistic value of 10^{-5} W/m^2 (*Bekerat et al.*, 2007). Figure 5.8 shows the GITM results driven by the small and large electron heat fluxes, with comparison to IRI and the Millstone Hill ISR measurements. The time shown is from 16-24 Local Time, which is from the dashed line to the end of the day in Figure 5.7. Before about 22 LT, the observed T_e was approximately 1000 K higher than GITM driven by the small heat flux above 400 km. The ion temperature was also underestimated by about 500 K above 400 km compared to IRI and the ISR data. T_e in IRI had a similar magnitude as observation above 400 km, but was about 1000 K warmer below.

Driven by the large heat flux, the dayside T_e increased significantly above 400 km, which significantly improved the comparison with the observation compared to the small-flux case. Moreover, T_i was raised to a similar magnitude as observed by the radar through the heat transfer between electrons and ions. This was especially true at high altitudes, where the collisional heating between electrons and ions is more important. Since a simple empirical heat flux model was used in this simulation, differences between GITM and ISR results can be expected. In order to simulate both T_e and T_i at mid-latitudes, a well-specified electron heat flux at the top is needed.

Finally, a comparison between GITM, IRI and the Sondrestrom ISR is shown in Figure 5.9. The heat flux at this location was a constant small value of 10^{-9} W/m^2 as shown in Figure 5.4. T_e in GITM was smaller than IRI and the ISR data, which could be caused by a few reasons: an underestimation in the driver of particle precipitation, or the fact that the top heat flux specified in the simulation was too low. T_i in GITM agreed pretty well with IRI: the T_i above 450 km was about 500 K higher than the T_i below. However, both GITM and IRI underestimated T_i around midnight. This could have been caused by not capturing the ion frictional heating term well, which could arise from a non-precise description of the ion or neutral flows in GITM (*Liuzzo et al.*,

2015). An underestimation of T_e could also lead to a lower value of T_i, since energy from the electrons flows to the ions through the temperature difference. Further study is needed to determine the high-latitude drivers of the ionospheric temperatures.

5.4 Results

5.4.1 Ionospheric temperature response to an idealized substorm

GITM with the improved electron and ion energy equations was used to investigate the ionospheric response to a idealized substorm, which was the same as Substorm 4 investigated by *Liu and Ridley* (2015). The prototypical substorm was constructed based on the superposed epoch variations of IMF B_z and HP during substorms using 5-years of Challenging Minisatellite Payload (CHAMP) (*Reigber et al.*, 2002) satellite data (*Clausen et al.*, 2014). The geomagnetic conditions used for the simulation are shown in Figure 5.10. B_z began to decrease at 1 hour before onset, reached a minimum of -3 nT at -25 min, and recovered to the initial value of zero at 80 minute. The background HP was 20 GW. It steadily increased after substorm onset, peaked at 55 GW at 20 min, and recovered to 20 GW at 2 hours epoch time.

In order to explore the effects of just the substorm, two simulations were conducted: one using the idealized substorm input as shown in Figure 5.10, termed Simulation 1, and the other using constant geomagnetic conditions with B_z=0 nT and HP=20 GW, termed Simulation 0. The response to the substorm was expressed as the percentage difference between the two simulations:

$$T\% = \frac{T_1 - T_0}{T_0} \times 100\%, \tag{5.25}$$

where T_1 is a specific ionospheric/thermospheric parameter in Simulation 1 and T_o is the parameter in Simulation 0. Figure 5.11 shows the ionospheric response to the substorm at 140 km geometrically averaged over the northern polar cap (i.e., above

50° Latitude). N_e had a immediate response to the HP variation due to the small time scale of the E region (i.e., in seconds). The response of T_e was highly correlated with the HP, with a cross correlation of 0.99, because the auroral ionization heating was the dominant source for T_e and the time scale of the E-region T_e modification was small (∼s). The percentage change of T_e peaked around 4% when the HP reached its maximum value. There was a small enhancement that appeared in T_e before the substorm onset. This was due to the ion-electron energy transfer during the growth phase of the substorm. Different than T_e, T_i started to respond during the growth phase of the substorm when the IMF B_z turned southward. The percentage difference of T_i peaked at 3% around the same time when B_z became most southward. The IMF-driven electric field was calculated as (*Kan and Lee*, 1979):

$$E_{IMF} = V_s B_s sin(\theta/2), \tag{5.26}$$

where V_s is the solar wind speed in the sunward direction, $B_s = ({B_y}^2 + {B_z}^2)^{\frac{1}{2}}$, and θ is the angle between the two components of the IMF, B_y and B_z. The response of T_i and the square of IMF electric field were highly correlated with a correlation coefficient of 0.97. This was because ions in the high-latitude E-region were in a quasi-energy balance by collision with neutrals (*Svenes et al.*, 1992a). With this assumption, T_i can be approximated as (*Schunk and Nagy*, 2009; *Thayer and Semeter*, 2004):

$$T_i = T_n + \frac{m_n}{3\kappa}(\boldsymbol{u_n} - \boldsymbol{u_i})^2. \tag{5.27}$$

The estimated T_i was dependent on the neutral temperature and the difference in bulk velocity between the ions and neutrals. The responding profiles of the estimated T_i and the simulated T_i had a high correlation of 0.99. The estimated T_i was always cooler than the simulated T_i because the energy balance assumption was not completely accurate for ions in the E region where electrons were sufficiently dense to

affect the ion temperature through heat transfer. The temporal profile of T_i variation was quite similar to the variation of δV^2, which is $(\boldsymbol{u_n} - \boldsymbol{u_i})^2$. This indicates that the velocity difference between the ions and neutrals, which represents the ion-neutral collisional heating, was the dominant heat source for ions during the substorm. As shown in the bottom panel in Figure 5.11, the variation of the velocity difference was controlled by the IMF-induced electric field. This was expected because the IMF electric field driver changed the ion drift significantly, while the neutral winds did not change much during the substorm (*St-Maurice and Schunk*, 1979). During the recovery phase of the substorm (i.e., about 60 min), the neutrals started to ramp up (*Liu and Ridley*, 2015) affecting the velocity difference and altering T_i.

Figure 5.12 shows the spatial variation of the ionospheric response at 6 times around the substorm onset. The leftmost column shows the background ionospheric state 60 min before the substorm onset, while the after plots show perturbations of the state variables with respect to Simulation 0. The perturbation in N_e was relatively small until after the substorm onset. N_e was enhanced by $\sim 100\%$ within the auroral oval when the HP reached maximum value. The red bands on the inner and outer boundaries were due to an expansion of the auroral oval and a low N_e background in these regions. Similar to N_e, T_e did not show noticeable perturbation until the HP increased when the auroral heating increased T_e by about 25% in the auroral oval. The perturbation in T_i started during the growth phase of the substorm as a response to the southward turning of IMF. The maximum enhancement occurred in the low-latitude convection zone, where large shears existed between the day-to-night neutral wind and the sunward ion flows. The perturbation died down before the onset time when B_z recovered.

Figure 5.13 shows the ionospheric response at 300 km in the same format as Figure 5.11. The response of N_e at 300 km was relatively sluggish compared to the response at 140 km. The cross correlation between HP and the N_e perturbation

peaked at 0.78 with a time delay of 30 minutes. This is because ∼keV precipitating electrons mainly deposited their energy in the E-region (*Gombosi*, 1998). Three factors could account for the enhancement of N_e in the F-region: (a) an increase of T_e led to a decrease in the dissociative recombination rate, which is discussed in detail below; (b) plasma was transported from the E-region due to diffusion along the magnetic field; (c) the lower energy precipitating particles in the distribution caused some ionization in the F-region. Further investigation is needed to determine how these processes combine to influence the response of the F-region. However, the relatively large chemical time scale of the ions in the F-region led to a long production and life time of the electron density perturbation in the F-region.

As shown in the second panel in Figure 5.13, the response of T_e was in phase with the variation of the HP, which suggested that a major source of T_e in the F region was heat conducted from the E region since there was little auroral heating in the F-region. The fact that the F-region temperature was highly correlated with the E-region density instead of the F-region density was also observed by *Maehlum et al.* (1984) and *Wahlund et al.* (1993) during an intense aurora observed by EISCAT. A slightly negative perturbation of T_e occurred between 1 hour and 4 hour epoch time, which probably because the heat conduction was into the F-region during the first part of the substorm and then out of the F-region during the second part of the substorm.

The estimate T_i was quite similar to the simulated T_i during the substorm, which indicates that the electron-ion heat transfer was negligible in the high-latitude F-region (*Schunk and Nagy*, 2009), and that the ion-neutral coupling was the dominant process. The major heat source to T_i was ion Joule heating and the major cooling term was heat transfer to neutrals. Before the substorm onset, the ion Joule heating driven by the IMF electric field primarily contributed to the enhancement of T_i. After 0 hour epoch time, the neutral temperature gradually increased, which reduced the

speed of the cooling process of T_i. A slight difference between the simulated and estimate T_i appeared between 0 and 1.5 hours because the electron-ion heat transfer was enhanced due to an increase in N_e and T_e after the substorm onset.

Figure 5.14 shows the spatial distribution of the F-region ionospheric perturbation in the same format as Figure 5.12. A prominent increase in N_e appeared after HP reached a maximum. At 60 min, a perturbation of about 100% existed around the auroral oval, as well as a bright band of about 300% on the outer boundary. The width of the perturbed region was about 4 times the width of the background auroral oval, similar to what was observed in the E-region. The F-region perturbation lasted for more than 2 hours due to relatively slow charge exchange and recombination rates. The response of T_e in the F-region was similar to the E-region. The percentage change in T_e reached about 30% in the main oval and around the high-latitude boundary of the auroral oval, and about 60% near the low-latitude boundary of the auroral oval when HP peaked at 20 min. The response of T_i maximized at -25 min when the IMF was most southward. The ion temperature then gradually died down after the peak.

5.4.2 Heat flux effect

In order to investigate the influence of the electron heat flux on the ionosphere, two simulations of the idealized substorm (as shown in Figure 5.10) were conducted, one with a small heat flux and one with a large heat flux during the substorm. The heat flux distributions were the same, but the large heat flux was one order of magnitude larger than the smaller heat flux, as shown in Figure 5.4. Figure 5.15 shows the ionospheric response to the two different heat fluxes. The increased heat flux was most effective on increasing T_e on the nightside in the mid-latitudes where N_e was approximately 100 times smaller than on the dayside. T_e was enhanced by $\sim$100% at maximum in the dusk and midnight regions. No obvious enhancement in T_e existed on the dayside. The increase in the heat flux also raised N_e and TEC, especially on

the nightside. The enhancement in the electron density reached ∼80% at maximum, but occurred in the region in which the density was very low.

Figure 5.16 shows vertical profiles of T_e (left) and N_e (right) at 150° Lon and 55° Lat at 1200 UT (i.e., 2200 LT), which is in the middle of the electron heat flux region on the nightside. The dotted and the solid lines represent the simulation results with the small and large heat fluxes, respectively. The increase in the heat flux raised T_e above approximately 140 km. The enhancement increased with altitude, and reached over 1000 K at 600 km. N_e was also increased above 140 km. The enhancement in N_e due to the increase in the heat flux was about $3 \times 10^9\ m^{-3}$ ($\sim 60\%$) above 230 km. This result agreed well with *David et al.* (2011).

The fact that the increase in the heat flux increased N_e was due to the combination of three factors in the ionosphere (*Schunk and Nagy*, 2009):

1. An increase in T_e can increase the ionospheric scale height, which reduces the exponential decrease of N_e with altitude above the F-region peak.

2. An increase in T_e can raise T_i through the electron-ion collisional heat transfer. The increase of T_i accelerates the charge exchange ratio between O^+ and molecular neutrals N_2, O_2 and NO. As the dissociative recombination rates of the resulting molecular ions are approximately five orders of magnitude faster than the radiative recombination rates of O^+, the increase in T_e could resulted in a decrease of N_e by raising T_i.

3. The radiative and dissociative recombination rates of atomic and molecular ion, respectively, are inversely proportional to T_e. Therefore, an increase of T_e could lead to an increase of N_e by reducing the recombination rates.

Here, a simple calculation was done to determine the net effect of changing either the electron or ion temperature on the F-region density due to the charge exchange

and dissociative recombination reactions. All the other ionospheric processes, such as transport and photoionization, were ignored.

Three charge exchange reactions were considered:

$$O^+ + N_2 \stackrel{\kappa_1}{\rightarrow} NO^+ + N, \tag{5.28}$$

$$O^+ + O_2 \stackrel{\kappa_2}{\rightarrow} O_2^+ + O, \tag{5.29}$$

$$O^+ + NO \stackrel{\kappa_3}{\rightarrow} NO^+ + O, \tag{5.30}$$

where the reaction rates κ_1, κ_2, and κ_3 are T_i dependent, and given in Appendix C. Two dissociative recombination reactions and the O^+ radiative recombination reaction were considered:

$$NO^+ + e^- \stackrel{\kappa_4}{\rightarrow} O + N, \tag{5.31}$$

$$O_2^+ + e^- \stackrel{\kappa_5}{\rightarrow} O + O, \tag{5.32}$$

$$O^+ + e^- \stackrel{\kappa_6}{\rightarrow} O, \tag{5.33}$$

where the reaction rates κ_4, κ_5, κ_6 are T_e dependent, and given in Appendix C. The initial ion densities (in m^{-3}) were set as: $n(O^+) = 1.0 \times 10^{12}$, $n(NO^+) = 0.0$ and $n(O_2^+) = 0.0$. The initial neural densities (in m^{-3}) were set as $n(NO) = 1.0 \times 10^{10}$, $n(O_2) = 1.0 \times 10^{11}$ and $n(N_2) = 1.0 \times 10^{12}$. The background $O+$ number density and neutral densities were typical F-region values from the simulation. Two steps were taken: (a) the charge exchange reactions were processed for one second, converting

O^+ to NO^+ and O_2^+; and (b) the recombination reactions were then processed for one second. The ion loss rate was obtained by subtracting the initial total ion density from the final total ion density. The calculation was conducted in the range of: T_e ~[1000 K, 6000 K] and T_i ~[1000 K, 3000 K]. The loss rate was normalized at the state of T_e = 1000 K and T_i = 1000 K, as given in Figure 5.17. The highest ion loss rate existed in the upper left corner (i.e., ion temperature much larger than electron temperature), where both the charge exchange rates were most enhanced but the dissociative recombination rates were most reduced. This could happened in regions with strong Joule heating, raising T_i even higher than T_e (*Schunk and Nagy*, 1978a; *Schunk et al.*, 1986a). A strong decrease in ion density would be expected in such a region. Likewise, the lowest ion loss rate occurred in the lower right corner (i.e., electron temperature much larger than ion temperature), where the charge exchange rates were mostly suppressed, but the dissociative recombination rates were enhanced. This circumstance can be expected in regions with strong downward electron heat flux (*David et al.*, 2011) or upward field-aligned currents, where T_e could be significantly increased (*Schunk et al.*, 1986a; *Zhang et al.*, 2003). The GITM simulation results shown in Figure 5.16 showed a positive relationship between T_e and N_e, which indicates that the effects of T_e on N_e caused by the change of the dissociative recombination rates (and the change of the ionospheric scale height) were dominant over the change of the charge exchange rates.

5.5 Summary and Conclusion

The improvement of the electron and ion temperature calculation in the Global Ionosphere Thermosphere Model has been described in this chapter. Time-dependent energy equations for T_e and T_i with heating and cooling rates, as well as field-aligned thermal conduction were implemented. The sources of energy in the electron equation include heating due to photoionization and ionization due to particle precipitation,

elastic collisions with ions, elastic and inelastic collisions with neutrals, as well as energy flux from the ring current/plasmasphere interaction. The sources of the ion temperature equation include elastic collisions with electrons and neutrals as well as frictional heating with neutrals.

The new GITM produced higher T_e and T_i at middle and high latitudes than the old model, which led to a denser ionosphere that was more comparable to the ionosphere in IRI. The new GITM reproduced the global structure of T_e observed in IRI including morning and evening enhancements at low and mid-latitudes. An expected anti-correlation between T_e and N_e was also observed in the new model. The temporal variations of the vertical distribution of T_e and T_i were similar to the Arecibo ISR observations. Compared to the observations by the Jicamarca ISR, GITM generally captured the magnitudes and the diurnal variations of T_e and T_i, but missed the small-scale structures on the nightside. It was also found that the mid-latitude T_e and T_i were strongly related to the electron heat flux. GITM needed a larger, more realistic heat flux input at the top boundary in order to reproduce similar magnitudes of T_e and T_i compared to the Millstone ISR observations. GITM also reproduced the general variations observed by the Sondrestrom ISR, but underestimated T_e and T_i, which was probably due to the lack of the heat flux in the auroral oval in the model and a lack of proper electric field and precipitation flux description to capture the realistic frictional heating.

The ionospheric response to an idealized substorm was investigated using the new version of GITM. The variation in N_e and T_e in the E-region were highly correlated with the variation in HP due to rapid auroral ionization processes and the short chemical time scale ($\sim$s) in this region. T_i began to respond during the growth phase of the substorm when the IMF B_z turned southward. The variation of T_i was mainly controlled by the IMF-driven electric field because the ion Joule heating rate was dependent on the ion velocity and the change of ion drift was mainly driven by the

IMF, assuming a relatively steady neutral wind pattern. The response of the F-region T_e was not correlated with the F-region N_e, but was correlated with the E-region T_e and N_e, which is consistent with observations by *Maehlum et al.* (1984) and *Wahlund et al.* (1993). This was because there was little auroral deposition in the F-region. The electron thermal conduction efficiently and rapidly transported heat from the E-region to the F-region, while the time scale of the electron density variations in the F-region was hours. The variations of the F-region T_i were also dominated by the IMF-driven electric field through the ion flow velocity and the frictional energy coupling with the neutrals.

The increase of the topside heat flux by one order of magnitude in the mid-latitudes raised T_e, N_e and TEC by approximately 120%, 80% and 80% (maximum), respectively in the dusk-to-midnight sector. The influence of the increased heat flux on the nightside penetrated down to about 140 km. The dayside ionosphere did not show significant changes in any of the variables in the increased heat flux region, due to the large electron density. The charge exchange and dissociative recombination reactions, which are T_i and T_e dependent respectively, are the two important chemical processes controlling the loss rate of the F-region ionosphere (*Schunk and Nagy*, 2009). High T_e and relatively low T_i tend to reduce the F-region ion loss rate, which can cause the F-region density in the heat flux region to reduce at a slower rate than it would with minimal heat flux.

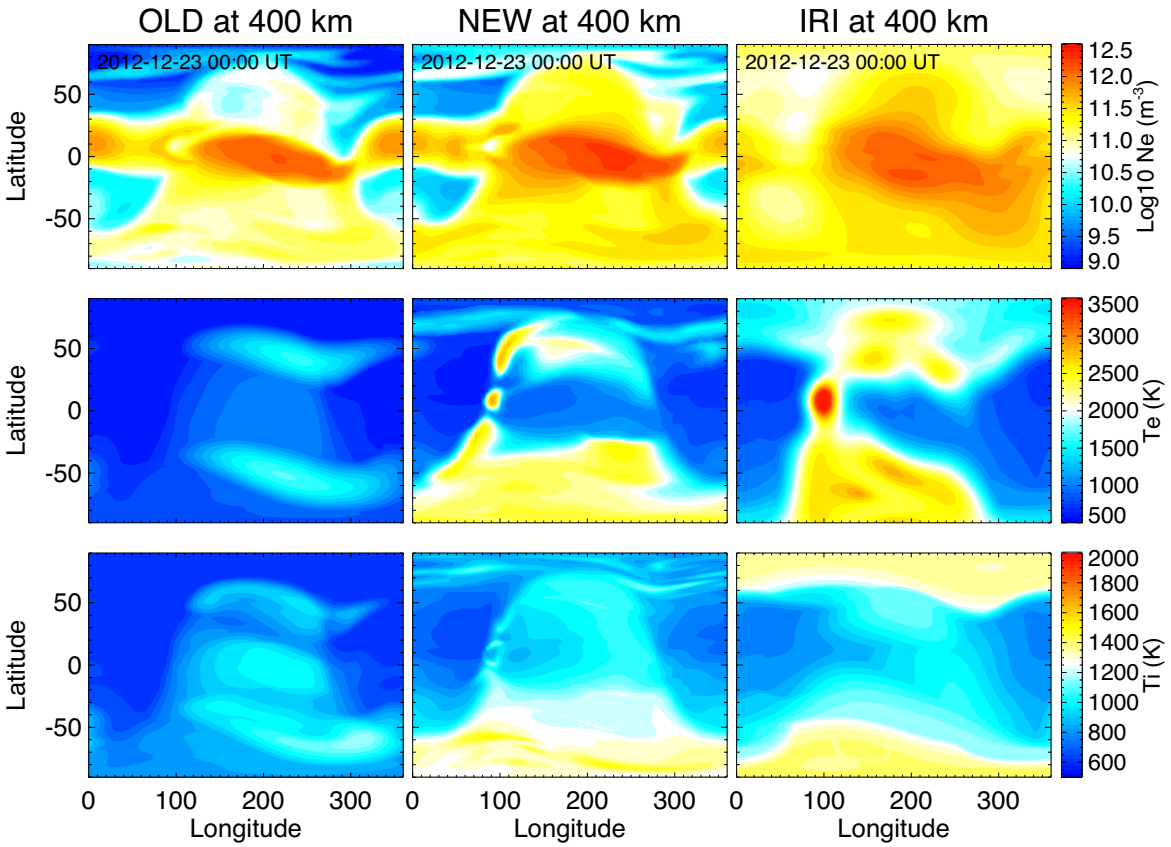

Figure 5.1: Comparisons of electron density (top), electron temperature (mid) and ion temperature (bottom) among the old GITM, the new GITM old and IRI at an altitudinal cut at 400 km at 00:00 UT on Nov 25, 2012.

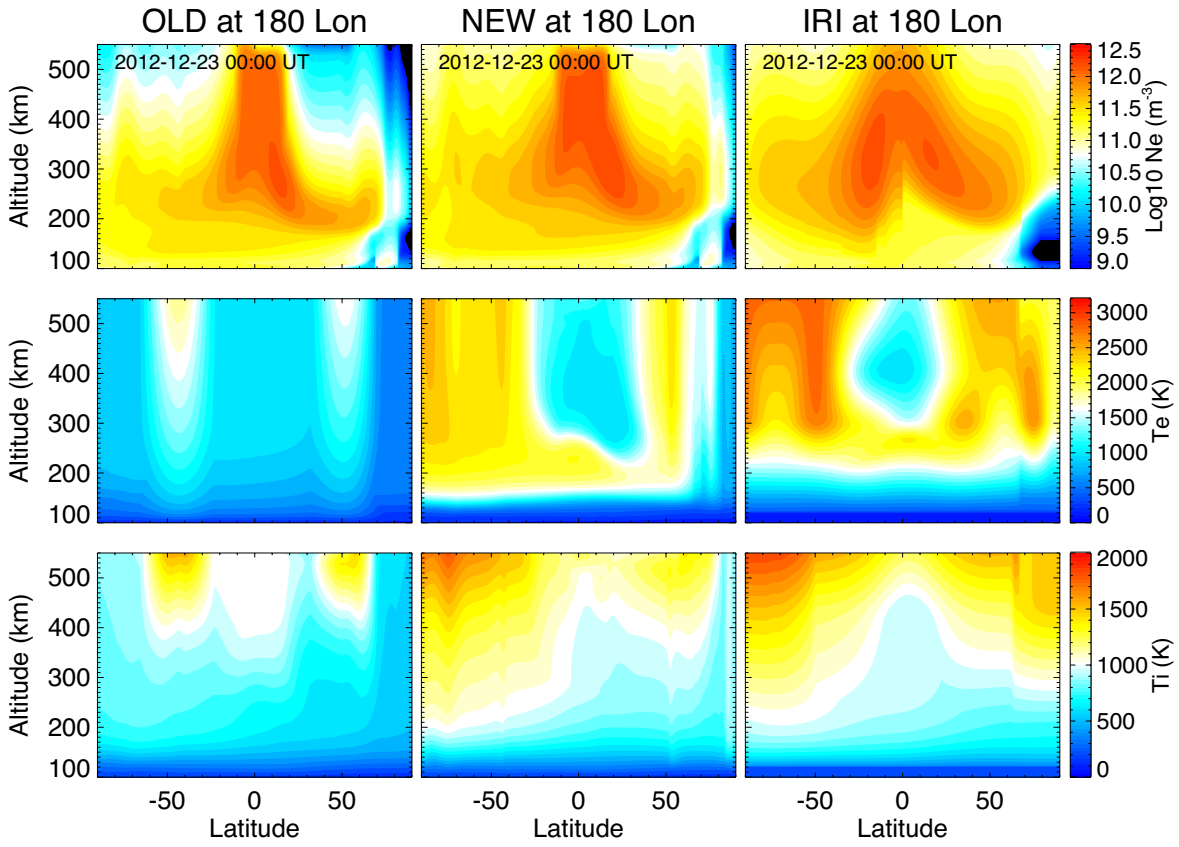

Figure 5.2: Comparisons of electron density (top), electron temperature (mid) and ion temperature (bottom) among the old GITM, the new GITM and IRI at a longitudinal cut at 180° at 00:00 UT on Nov 25, 2012.

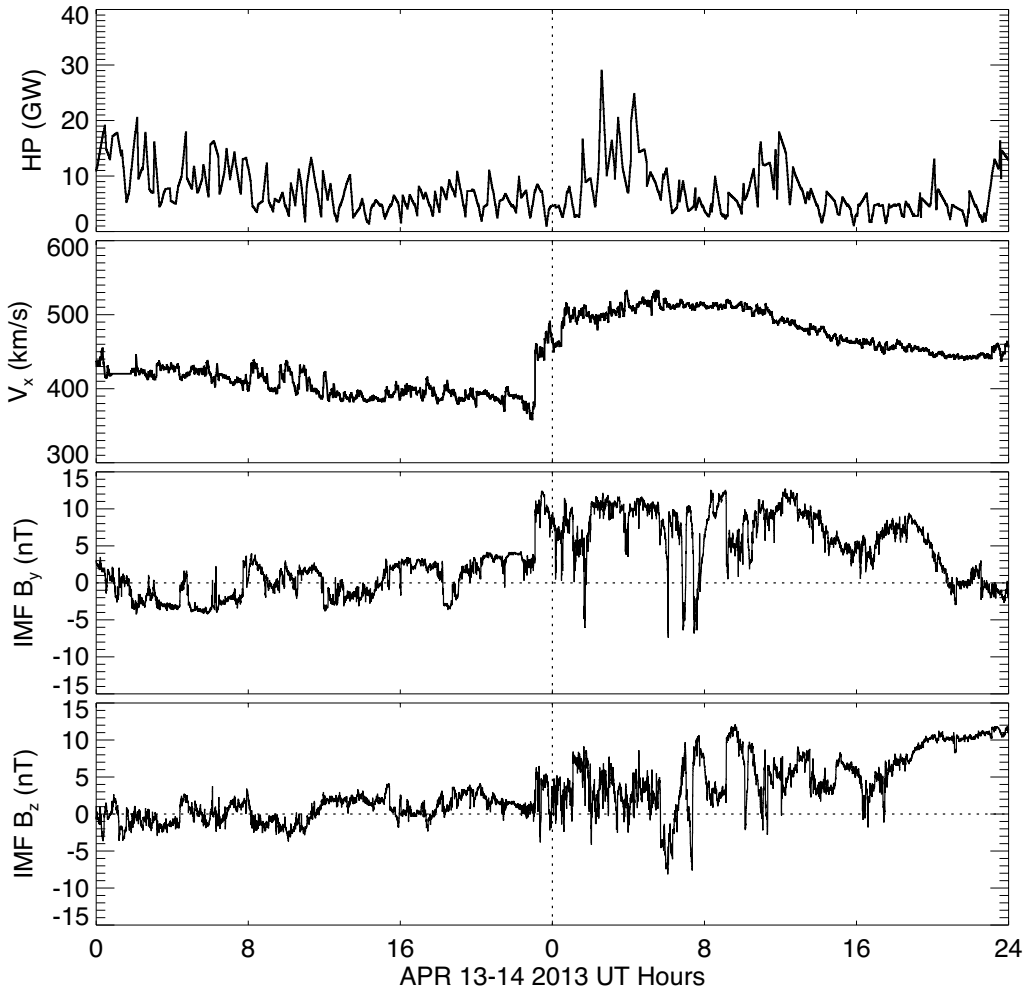

Figure 5.3: Temporal variations of hemispheric power, solar wind velocity Vx, IMF By and IMF Bz in GSE coordinates from top to bottom on April 13-14, 2013.

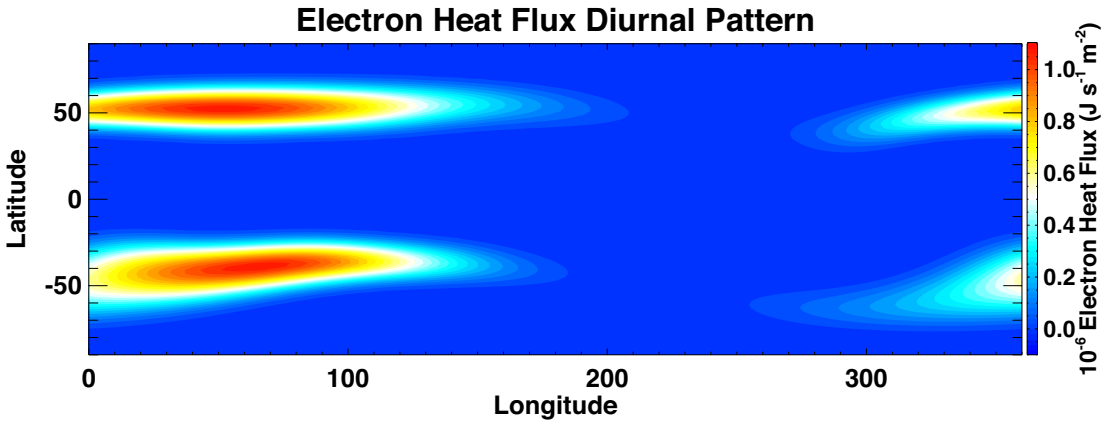

Figure 5.4: Global distribution of the electron heating flux at the top boundary at 12:00 UT.

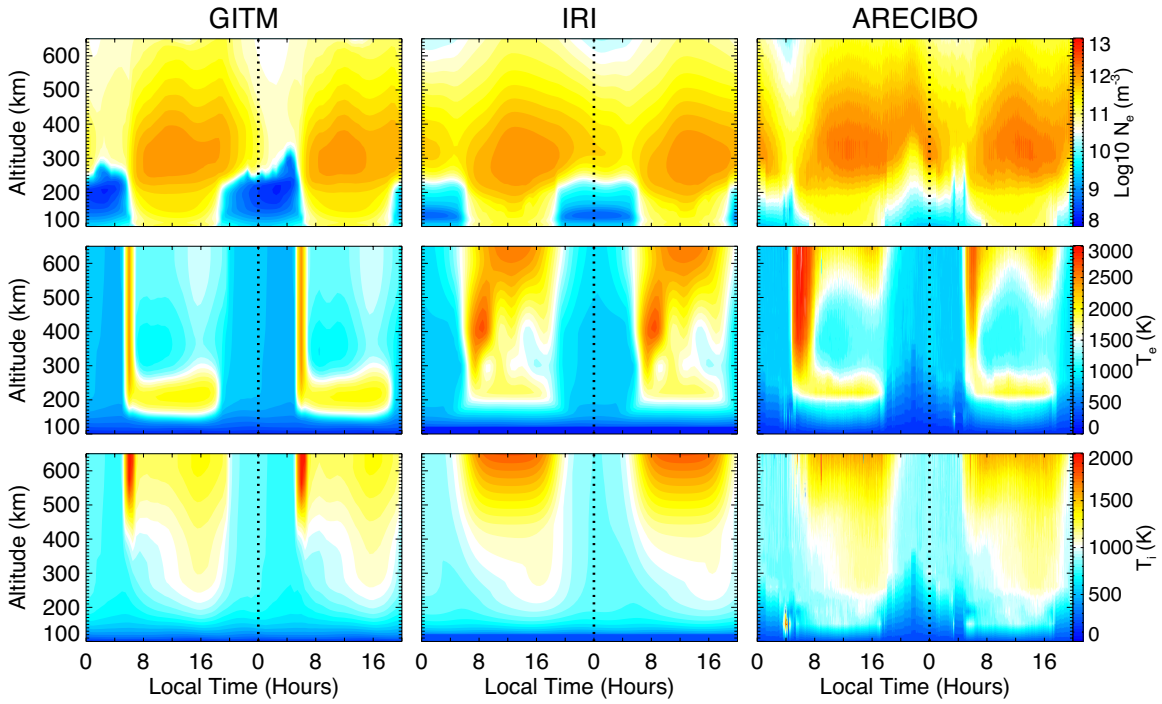

Figure 5.5: Comparisons of temporal variations in the electron density (top), electron temperature (mid) and ion temperature (bottom) between GITM (left column), IRI (right column) and observation by the Arecibo ISR (right column) from 100 km to 650 km on April 13th and 14th, 2013.

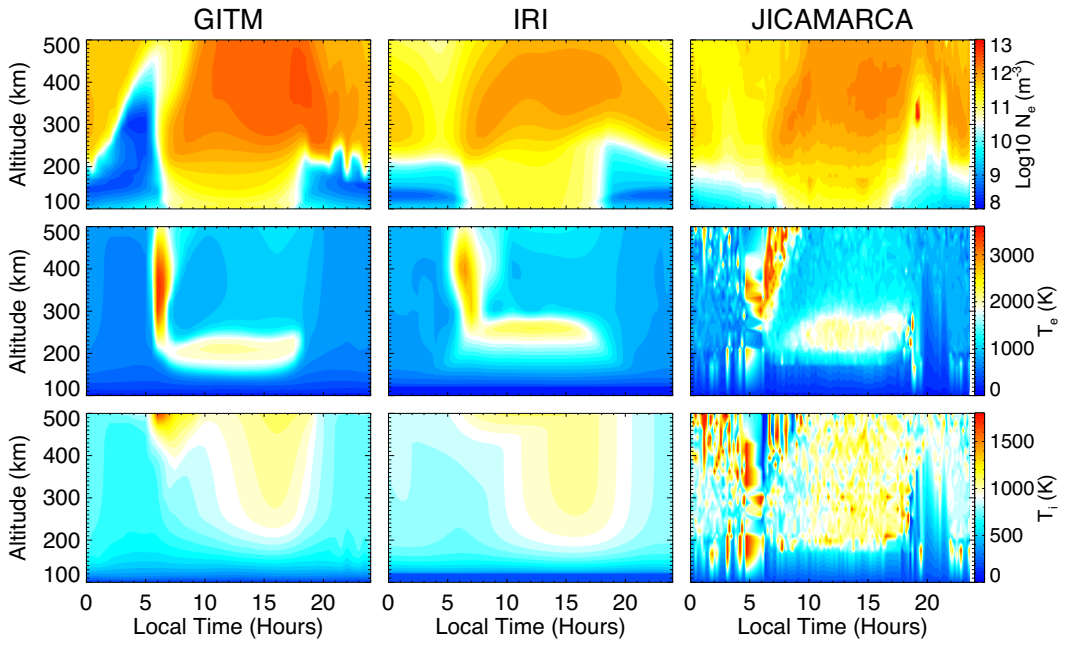

Figure 5.6: Comparisons of temporal variations in the electron density (top), electron temperature (mid) and ion temperature (bottom) between GITM (left column), IRI (center column) and observation by the Jicamarca ISR (right column) from 100 km to 650 km on April 13th, 2013.

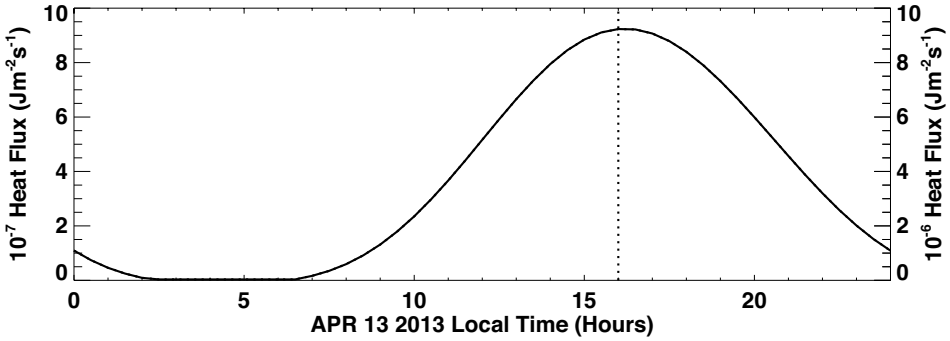

Figure 5.7: The electron heat flux specified on the topside of the ionosphere at Millstone ISR from 00 to 24 LT on April 13th, 2013. The left (right) Y-axis shows the magnitude of the small (large) heat flux.The dashed line marks 16 LT.

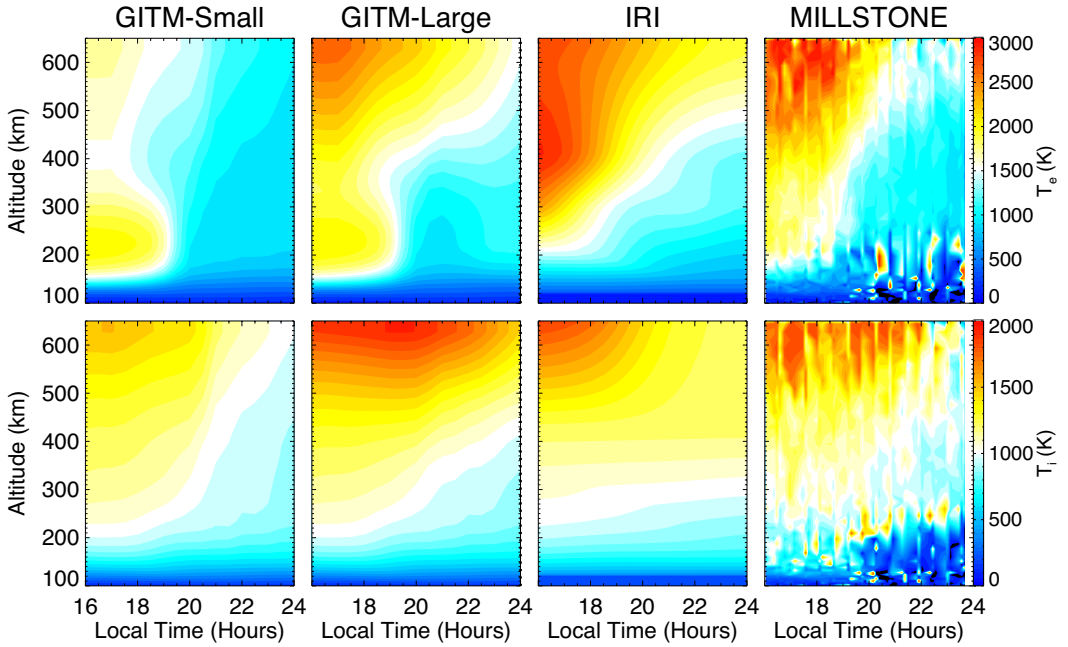

Figure 5.8: Comparisons of temporal variations in the electron temperature (top) and ion temperature (bottom) between the simulations (left three columns) and the Millstone observation (right column) from 100 km to 650 km from 16 to 24 LT on April 13th, 2013. The leftmost column shows T_e and T_i of the simulation driven by the small heat flux, while the second column from the left shows T_e and T_i of the simulation driven by the large heat flux. The third column column for the left shows results from IRI.

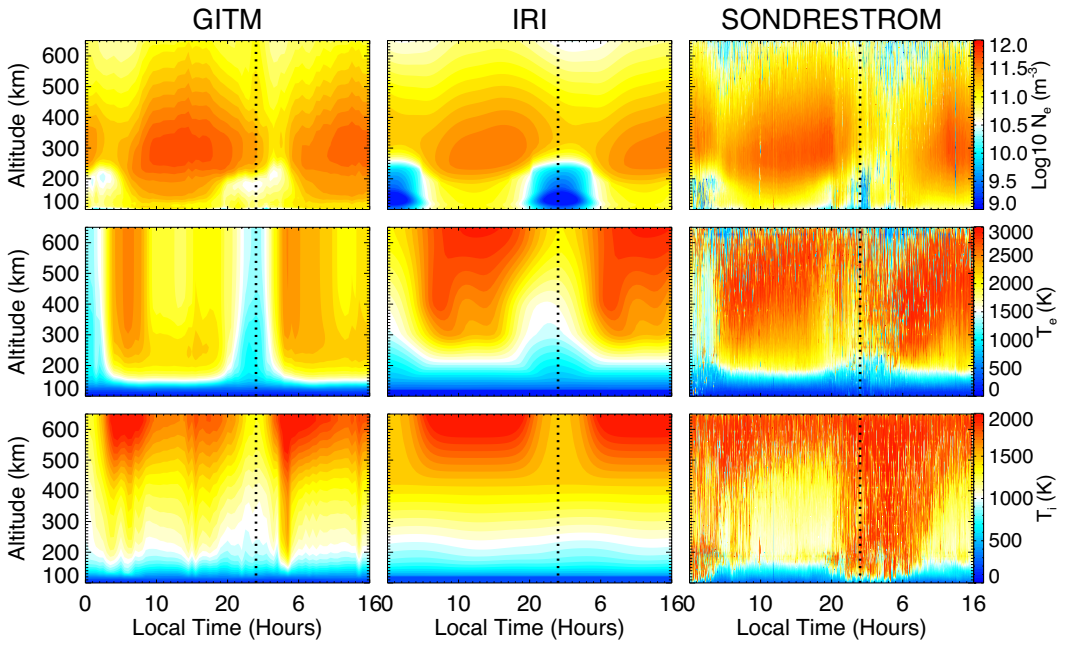

Figure 5.9: Comparisons of temporal variations in the electron density (top), electron temperature (mid) and ion temperature (bottom) between GITM (left column), IRI (center column) and observation by the Sondrestrom ISR (right column) from 100 km to 650 km on April 13th and 14th, 2013.

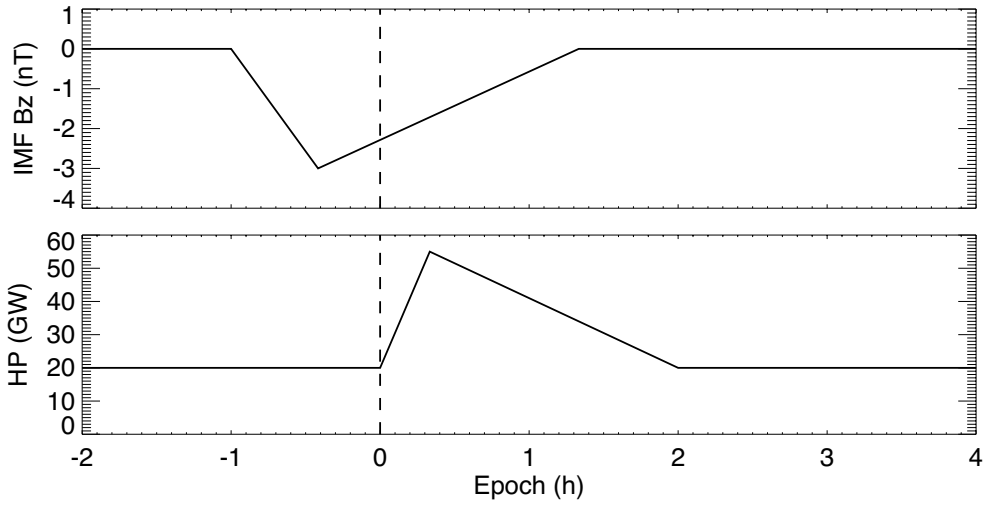

Figure 5.10: Temporal variations of IMF B_Z and hemispheric power (HP) during an idealized substorm. The onset occurred at 0 epoch time, as indicated as the vertical dashed line.

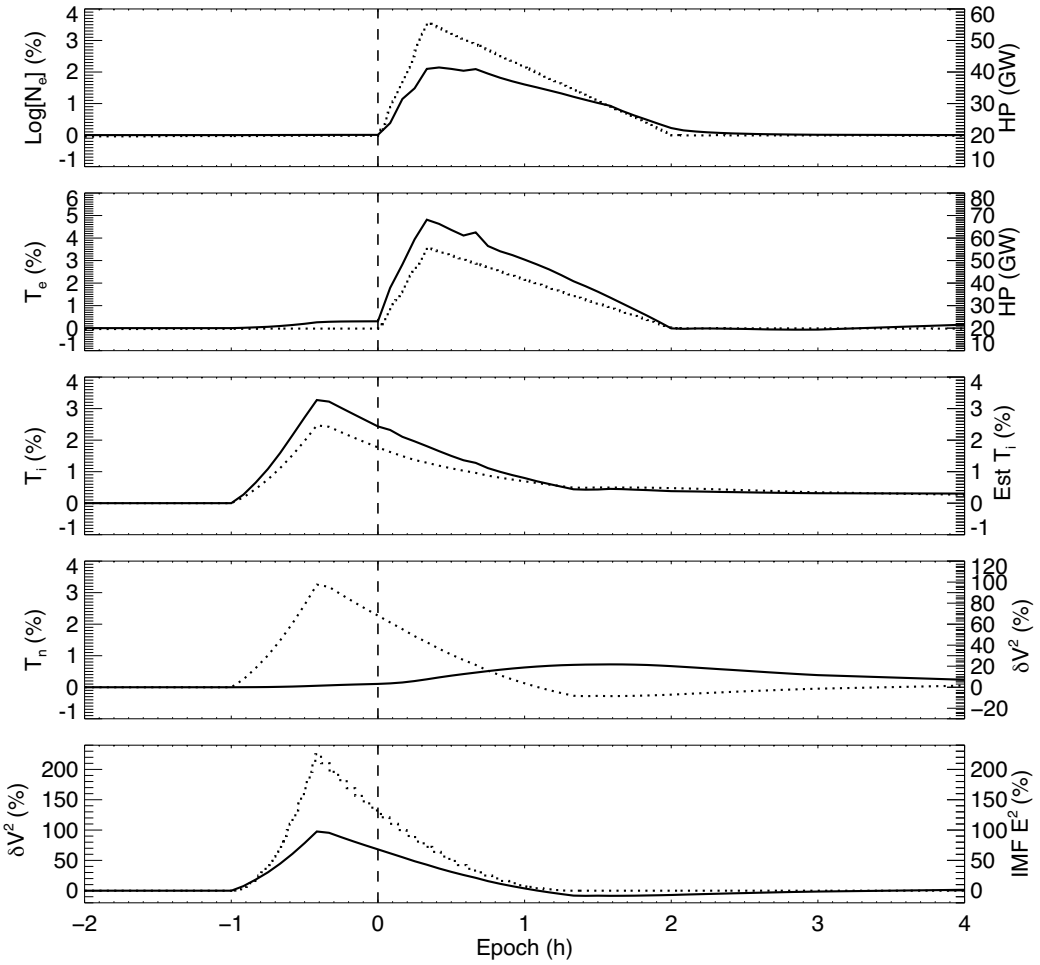

Figure 5.11: Solid lines show from top to bottom temporal variations of percentage difference in N_e, T_e, T_i, T_n and square of velocity difference between ions and neutrals (δV^2); dotted lines show from top to bottom HP, the percentage difference in estimated T_i, δV^2 and square of IMF driven electric field. The ionospheric and thermospheric parameters are geometrically averaged over the north polar cap (above 40° Latitude) at 140 km.

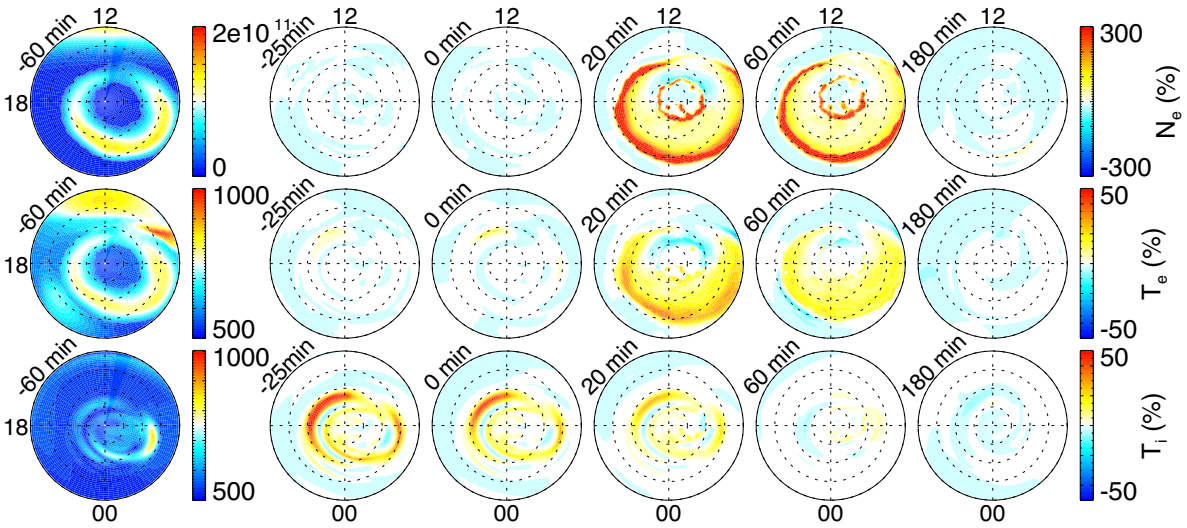

Figure 5.12: From top to bottom: spatial variations of N_e, T_e and T_i at 140-km altitude at 6 epoch times during the substorm. The leftmost column shows the background ionosphere one hour before the substorm onset. Columns 2-6 show the percentage difference between the substorm simulation and the background simulation at -25 min, 0 min, 20 min, 60 min and 180 min epoch times, respectively.

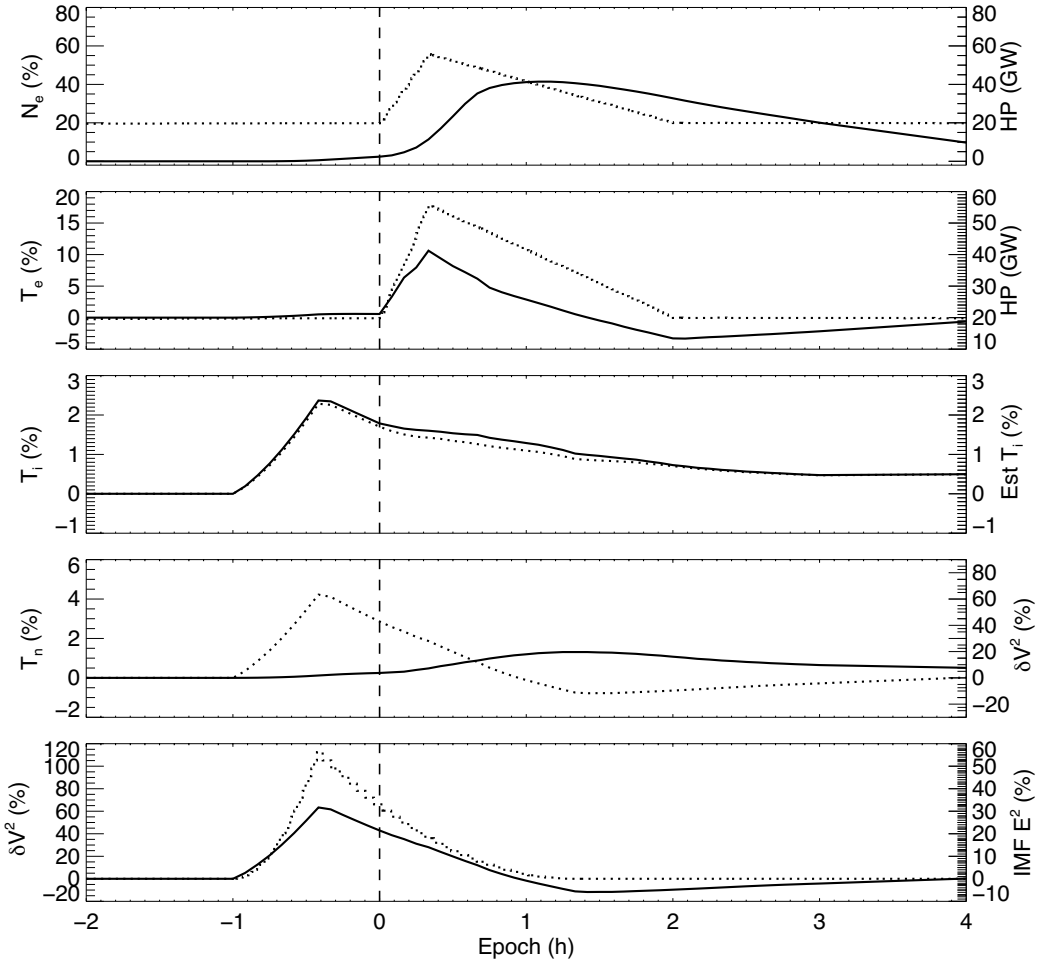

Figure 5.13: Same as Figure 5.11 but at 300 km.

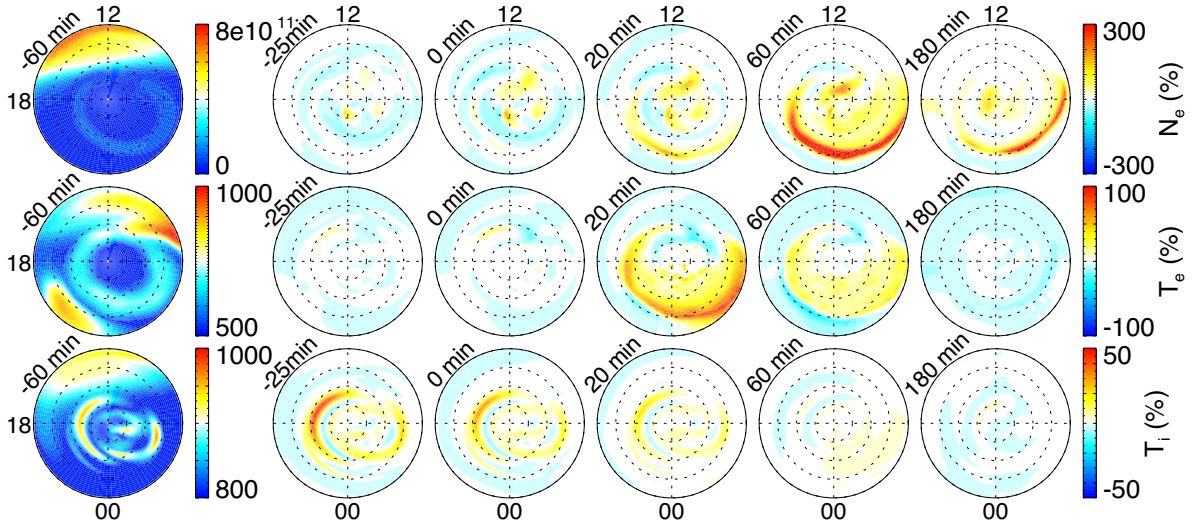

Figure 5.14: Same as Figure 5.12 but at 300 km.

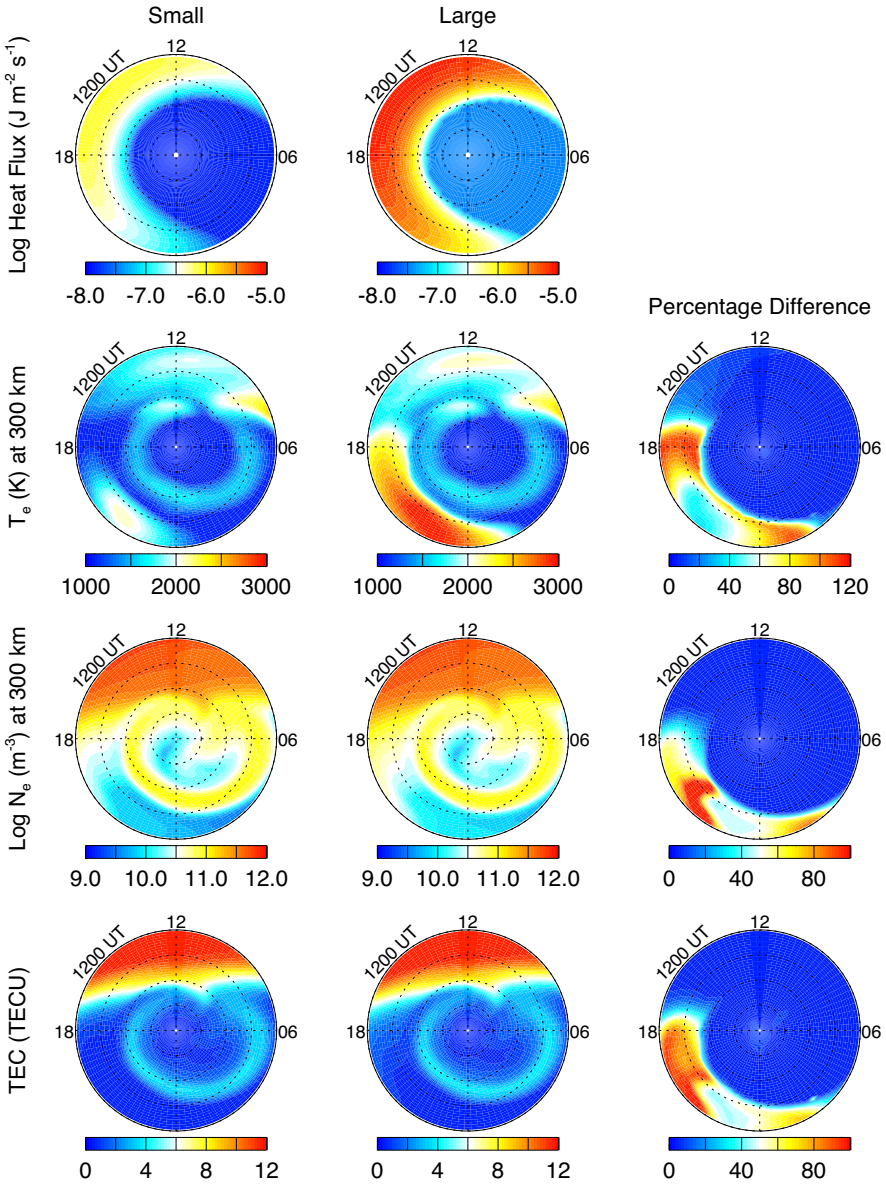

Figure 5.15: From top to bottom shows the electron heat flux into the ionosphere, T_e and N_e at an altitude of 300 km and the total electron content (TEC) at 12:00 UT . The first (second) column shows the simulation results driven by the small (large) heat flux, and the third column shows the percentage difference between the two.

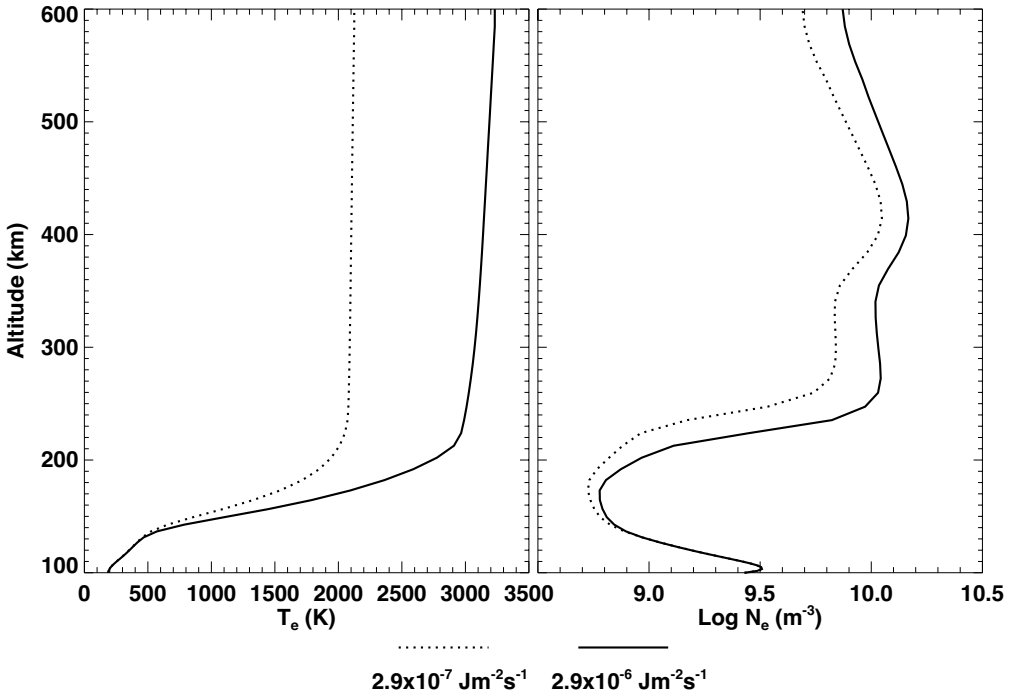

Figure 5.16: Altitudinal profiles of the electron temperature and electron density at 150° Lon 55° Lat at 12:00 UT (i.e., 22:00 LT). The dotted lines represent the top heat flux of $2.9\times10^{-7}Jm^{-2}s^{-1}$, while the solid lines represent the top heat flux of $2.9\times10^{-6}Jm^{-2}s^{-1}$.

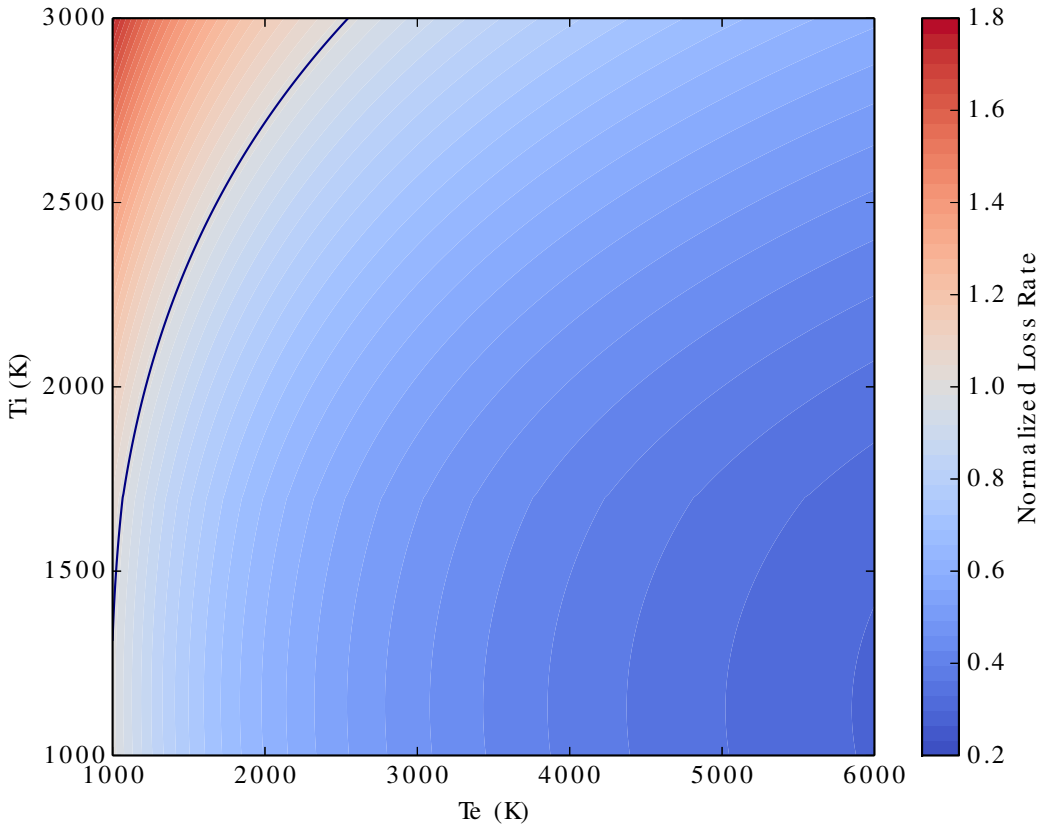

Figure 5.17: The electron density loss rate normalized to the value at Te=1000 K, T_i=1000 K. Level = 1.0 is contoured by the solid line.

CHAPTER VI

Investigating the performance of simplified neutral-ion collisional heating rate in a global IT model

The Joule heating rate has usually been used as an approximate form of the neutral-ion collisional heating rate in the thermospheric energy equation in global thermosphere-ionosphere models. This means that the energy coupling has ignored the energy gained by the ions from collisions with electrons. It was found that the globally averaged thermospheric temperature (T_n) was underestimated in simulations using the Joule heating rate, by about 11% when $F10.7$=110 sfu in a quiet geomagnetic condition. The underestimation of T_n was higher at low latitudes than high latitudes, and higher at F-region altitudes than at E-region altitudes. It was found that adding additional neutral photoelectron heating in a global IT model compensated for the underestimation of T_n using the Joule heating approximation. Adding direct photoelectron heating to the neutrals compensated for the indirect path for the energy that flows from the electrons to the ions then to the neutrals naturally, and therefore was an adequate compensation over the dayside. There was a slight dependence of the underestimation of T_n on $F10.7$, such that larger activity levels resulted in a need for more compensation in direct photoelectron heating to the neutrals to

make up for the neglected indirect heating through ions and electrons.

6.1 Introduction

The energy coupling between the ionospheric plasma and the neutral atmosphere strongly affects the global energy budget and temperature distribution of the thermosphere. Ionosphere-thermosphere models usually use the Joule heating approximation as the neutral-ionneutral-ion energy coupling term in the neutral energy equation (*Roble et al.*, 1988; *Fuller-Rowell and Rees*, 1980; *Zhu et al.*, 2005).Various studies have shown that Joule heating (*Cole*, 1962, 1971) is one of the major energy sources of the upper atmosphere at high latitudes using satellite (*Heelis and Coley*, 1988; *Gary et al.*, 1995; *Liu and Lühr*, 2005) and ground-based measurements (*Banks et al.*, 1981; *Thayer*, 1998), as well as using coupled global ionosphere thermosphere models (*Barth et al.*, 2009; *Fuller-Rowell et al.*, 1996; *Rodger et al.*, 2001; *Deng et al.*, 2011). *Codrescu et al.* (1995) suggested that Joule heating could be significantly underestimated by the exclusion of small-scale variability of E-field in high-latitude convection models. *Deng and Ridley* (2007) further pointed out that model resolution and the vertical differences between ion and neutral velocity were other two sources for an underestimation of Joule heating within global IT models. *Emery et al.* (1999) suggested that a corrective multiplicative factor of 2.5 of the Joule heating rate was needed for the winter hemisphere in order to account for small scale structures and rapid variability in high-latitude electric fields. Significant efforts have been made to quantify various uncertainties existing in modeling the Joule heating rate. However, despite the widespread use of the Joule heating rate as an approximation of the neutral-ion collisional heating rate in the neutral energy equation, there have been few studies showing how well the Joule heating rate performs in a global ionosphere thermosphere model.

Solar radiation in Extreme Ultraviolet (EUV) and soft X-ray wavelengths is the

dominant energy source of the upper atmosphere. It is known that the solar radiation energy primarily goes directly into photoionization and molecular dissociation (*Torr et al.*, 1980a). Photoelectrons are produced through the photoionization process, carrying photon energy in excess of the ionization threshold as kinetic energy. Photoelectrons are then responsible for heating the ambient thermal electrons (*Smithtro and Solomon*, 2008). Efforts have been made to develop a physical model to solve photoelectron flux and energy spectra considering transport, elastic and inelastic collisions, and energy loss to ambient electrons (*Nagy and Banks*, 1970; *Richards and Torr*, 1983; *Torr et al.*, 1990). A parameterization of the electron volume heating rate by photoelectrons was developed by *Swartz and Nisbet* (1972). An improved parameterization was further developed for application to photoelectron heating of ambient thermal electrons during solar flares (*Smithtro and Solomon*, 2008). It was suggested that neutrals were indirectly heated by photoelectrons: collisions between photoelectrons and thermal electrons produced hot electrons which heat neutrals and ions through elastic and inelastic collisions (*Torr et al.*, 1980a; *Roble and Emery*, 1983; *Aggarwal et al.*, 1979). The electron-ion collisions dominate over the electron-neutral collisions above the E-region (*Aggarwal et al.*, 1979). A constant photoelectron heating efficiency of $\sim$0.05 has been applied in global ionosphere thermosphere models such as the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIEGCM) (*Roble et al.*, 1988; *Richmond*, 1995) and the Global Ionosphere Thermosphere Model (GITM) (*Ridley et al.*, 2006) in order to compensate for the discrepancy in the thermospheric temperature between model results and observations (*Burrell et al.*, 2015; *Maute*, 2011). However, there have been few literatures investigating whether there exists direct photoelectron heating to the neutral atmosphere, or quantifying the neutral photoelectron heating efficiency for the neutral atmosphere either by observation or by numerical calculation to the author's knowledge.

In this study, through the investigation of the performance of the Joule heating

rate as an approximate form of the neutral-ion energy coupling rate in GITM, an explanation (or a justification) for using a photoelectron heating efficiency for the neutral atmosphere will be presented. To fully consider the neutral-ion energy coupling, a complete neutral-ion collisional heating terms need to be considered. Two forms of the neutral-ion heating rate were implemented in GITM: the simplified Joule heating rate and a more complete energy equation that allows energy flow from the electrons to the ions then to the neutrals. The influence of the two forms of neutral-ion heating rate on the thermospheric temperature was investigated and three questions will be explored: (a) How much has T_n been underestimated or overestimated by using the Joule heating as the neutral-ion energy coupling term? (b) How did the performance of the Joule heating term change with altitude, latitude, and local time? (c) How accurately has the neutral photoelectron heating used in global IT models compensated for the missing heating?

This study used the recently updated GITM (as described in Chapter V), in which the neutral, ion and electron energies are fully coupled.

The complete neutral-ion collisional heating rate can be written as (*Banks and Kockarts*, 1973; *Schunk*, 1975):

$$Q_C = \sum_k n_k m_k \sum_t \frac{\nu_{kt}}{m_k + m_t}[3\kappa(T_i - T_n) + m_t(\boldsymbol{u_n} - \boldsymbol{u_i})^2], \tag{6.1}$$

where n, m and T are the number density, mass and temperature respectively, $\boldsymbol{u_n}$ and $\boldsymbol{u_i}$ are the neutral and ion velocities, and the subscripts t and k denote the ion and neutral species, respectively, while the subscripts i and n denote the bulk ion and neutrals, respectively. Specifically, the term "neutral-ion" was used for source terms in the neutral energy equation and the term "ion-neutral" was used in the ion energy equation here. The first term is the heat transfer rate from the ions to the neutrals, with the second term being the neutral-ion frictional heating rate due to the velocity difference between the two species (*Banks and Kockarts*, 1973; *Schunk*, 1975).

Generally, the ion temperature can be assumed to be in steady-state, and balanced by energy coupling to both neutrals and electrons:

$$3\kappa(T_i - T_n) = m_n(\boldsymbol{u_n} - \boldsymbol{u_i})^2 + \frac{m_i + m_n}{m_i}\frac{\nu_{ie}}{\nu_{in}}(3\kappa(T_e - T_i) + m_e(\boldsymbol{u_e} - \boldsymbol{u_i})^2), \quad (6.2)$$

where ν_{ie} and ν_{in} are the collisional frequencies between ions and electrons and between ions and neutrals, respectively. Considering $m_e \ll m_i$, the ion-electron frictional heating rate can be ignored, and Equation 6.2 can be simplified to:

$$3\kappa(T_i - T_n) = m_n(\boldsymbol{u_n} - \boldsymbol{u_i})^2 + \frac{m_i + m_n}{m_i}\frac{\nu_{ie}}{\nu_{in}}3\kappa(T_e - T_i). \quad (6.3)$$

At high latitudes or on the nightside when the electron density is low, ν_{ie} can be much less than ν_{in}. Thus, a balance can be approximated between the ion-neutral heat transfer rate and the ion-neutral frictional heating rate, and the ion energy equation can be simplified to:

$$3\kappa(T_i - T_n) \cong m_n(\boldsymbol{u_n} - \boldsymbol{u_i})^2. \quad (6.4)$$

This assumption has been widely applied for large temporal and spatial ionospheric structure at high latitudes when the ion density is low (*St-Maurice and Hanson*, 1982; *Killeen et al.*, 1984; *Schunk and Nagy*, 2009; *Thayer and Semeter*, 2004).

This approximation can be substituted into Equation 6.1, so that the complete neutral-ion collisional heating rate can be written as:

$$Q_C \approx Q_J = \sum_k n_k m_k \sum_t \frac{\nu_{kt}}{m_k + m_t}[m_k(\boldsymbol{u_n} - \boldsymbol{u_i})^2 + m_t(\boldsymbol{u_n} - \boldsymbol{u_i})^2]. \quad (6.5)$$

This is consistent with the suggestion by *St-Maurice and Hanson* (1982) that the Joule heating rate was twice the neutral-ion frictional heating assuming $m_k \approx m_t$.

This equivalence was confirmed by in Situ measurements by the Atmosphere Explorer satellites around the 1980s (*St-Maurice and Hanson*, 1982).

Using the relation:

$$n_n m_n \nu_{ni} = n_i m_i \nu_{in}, \tag{6.6}$$

the neutral-ion collisional heating rate in Equation 6.5 can be written as:

$$Q_J = \sum_t n_t m_t \sum_k \nu_{tk} (\boldsymbol{u_n} - \boldsymbol{u_i})^2. \tag{6.7}$$

If the ion and electron motion perpendicular to the magnetic field is in steady state and determined only by the Lorentz and ion drag force, the electron gyro-frequency is much greater than the electron-neutral collisional frequency which is true above 90 km (*Brekke*, 2012; *Thayer and Semeter*, 2004; *Strangeway*, 2012), Q_J is equivalent to the Joule heating rate:

$$Q_J = \boldsymbol{j} \cdot \boldsymbol{E'}, \tag{6.8}$$

Here, $\boldsymbol{j}$ is current and $\boldsymbol{E'}$ is the electric field in the neutral gas frame (*Thayer and Semeter*, 2004).

The errors in the temporal change rate of T_n using Joule heating rate can be estimated by subtracting the time rate of change of T_n due to Q_J from that due to Q_C. The time rate of change of the neutral temperature due to the neutral-ion energy coupling is given by:

$$\frac{dT_n}{dt} = \frac{Q}{\kappa \sum_k n_k m_k}, \tag{6.9}$$

where κ is the boltzmann constant, and Q stands for either the Complete neutral-ion collisional heating rate as shown in Equation 6.1 or the Joule heating rate as shown

in Equation 6.7. For simplicity, the mean mass (i.e., number density weighted mass) was applied for the neutrals ($\bar{m_n}$) and ions ($\bar{m_i}$) in the following calculation. In order to explain the errors in the simplification in going from the more-complete equation to the Joule heating simplification, the difference between the two can be explored and expressed as:

$$\Delta \frac{dT_n}{dt} \propto \frac{Q_C - Q_J}{n_n \bar{m_n} \kappa}. \tag{6.10}$$

Q_C in Equation 6.1 was simplified to:

$$Q_{C_s} = n_n \bar{m_n} \frac{\nu_{ni}}{\bar{m_n} + \bar{m_i}} [3\kappa(T_i - T_n) + \bar{m_i}(\boldsymbol{u_n} - \boldsymbol{u_i})^2], \tag{6.11}$$

while Q_J in Equation 6.5 was simplified to:

$$Q_{J_s} = n_n \bar{m_n} \frac{\nu_{ni}}{\bar{m_n} + \bar{m_i}} [\bar{m_n}(\boldsymbol{u_n} - \boldsymbol{u_i})^2 + \bar{m_i}(\boldsymbol{u_n} - \boldsymbol{u_i})^2] \tag{6.12}$$

Substituting Q_{C_s} and Q_{J_s} into Equation 6.10 leads to:

$$\Delta \frac{dT_n}{dt} \propto \frac{\nu_{ni}}{\kappa(\bar{m_n} + \bar{m_i})} [3\kappa(T_i - T_n) - \bar{m_n}(\boldsymbol{u_n} - \boldsymbol{u_i})^2], \tag{6.13}$$

Substituting Equation 6.3 into Equation 6.13 and assuming $\bar{m_n} \approx \bar{m_i}$ results in:

$$\Delta \frac{dT_n}{dt} \propto \frac{\nu_{ie}}{\bar{m_i}} 3(T_e - T_i). \tag{6.14}$$

Considering the electron-ion collision frequency in s^{-1} (*Schunk and Nagy*, 2009):

$$\nu_{ei} = 5.44 \times 10^{-5} \frac{n_i Z_i^2}{T_e^{3/2}}, \tag{6.15}$$

where n_i is the ion number density in m^{-3}, Z_i is the number of ion charge, T_e is in K and 54.4 is in $\mathrm{s}^{-1}\ \mathrm{K}^{\frac{3}{2}}\ \mathrm{m}^3$, and the relation:

$$n_e m_e \nu_{ei} = n_i m_i \nu_{ie} \tag{6.16}$$

is used. Equation 6.14 can be further expressed as:

$$\Delta \frac{dT_n}{dt} \propto \frac{5.44 \times 10^{-5} Z_i^2 m_e}{\bar{m_i}^2} \frac{n_i}{T_e^{\frac{3}{2}}} (T_e - T_i) \propto \frac{n_e}{T_e^{\frac{3}{2}}} (T_e - T_i). \tag{6.17}$$

Here, the variation of the neutral-ion collisional frequency (which is neutral-density dependent) was ignored. This equation shows that the Joule heating approximation is valid in regions in which either (a) the ion density is quite low ; or (b) the temperature difference between the ions and electrons is small. As noted by *St-Maurice and Hanson* (1982); *Killeen et al.* (1984); *Schunk and Nagy* (2009); *Thayer and Semeter* (2004), these conditions tend to occur at high latitudes.

The heating rates are defined for reference below. The ion-neutral frictional heating rate is

$$Q_F(I-N) = \sum_t n_t m_t \sum_k \frac{m_k \nu_{tk}}{m_t + m_k} (\boldsymbol{u_n} - \boldsymbol{u_i})^2, \tag{6.18}$$

and the ion-neutral heat transfer rate is

$$Q_T(I-N) = \sum_t n_t m_t \sum_k \frac{3\kappa \nu_{tk}}{m_t + m_k} (T_i - T_n). \tag{6.19}$$

The ion-electron heat transfer rate is expressed as:

$$Q_T(I-E) = \sum_t n_t m_t \frac{3\kappa \nu_{te}}{m_t + m_e} (T_i - T_e). \tag{6.20}$$

In this study, the complete neutral-ion collisional heating rate in Equation 6.1 and the Joule heating rate in Equation 6.7 were implemented in the Global Ionosphere Thermosphere Model. First, a set of simulations were conducted during the winter solstice, i.e., Dec 21-23, 2012. The first simulation used a complete neutral-

ion collisional heating rate with zero photoelectron heating efficiency (PHE) (termed the Complete simulation). The second simulation used the Joule heating rate as an approximate form of the neutral-ion collisional heating rate with zero PHE (termed the Joule simulation). The third simulation (termed the Joule simulation with 0.05 PHE), is the same as the Joule simulation except with a PHE efficiency of 0.05. All external drivers in the three simulations were the same and constant: the $F10.7$ index was 110 sfu, IMF B_z was southward with the value of -2 nT and the IMF B_y was zero nT, the solar wind speed was 400 km/s, and the hemispheric power was set to 20 GW. The dynamo solver was turned on in all the simulations. The grid size was 2.5° longitude by 1.0° latitude. The altitudinal grid size was stretched to about $\frac{1}{3}$ of a scale height based on the initial thermospheric temperature and density. Also, the same set of simulations were conducted with two different $F10.7$: 70 sfu and 150 sfu, in order to explore the dependence of the Joule heating approximation on solar irradiance and justify the photoelectron heating efficiency used to compensate the missing heating.

6.2 Results

6.2.1 Comparison between the Complete and Joule simulations

Figure 6.1 shows the temporal variation of the globally volume-averaged neutral temperature over three simulation days for the three cases. The globally averaged temperature was plotted to illustrate the evolution of the simulations into a steady state. In all three cases, the globally averaged T_n decreased quickly during the first 5 hours, increased gradually and leveled off at the beginning of the third day. T_n dropped in the beginning of the simulation because GITM does not assume a hydrostatic solution (*Ridley et al.*, 2006), and it is initialized with MSIS (*Hedin et al.*, 1977), which does not have a perfectly hydrostatic balance. There were massive mod-

ifications of the dynamics that took place over the beginning of the simulations. T_n in the Joule simulation leveled off at around 740 K, which was about 90 K ($\sim$11%) lower than the Complete simulation. This was expected because the Joule heating rate did not account for the ion-electron heat transfer in the ion thermal equation. This resulted in an underestimation of $T_i - T_n$ as shown in Equation 6.4, thus leading to a lower neutral-ion heat transfer rate for the neutrals. This phenomenon tends to occur in the dayside F-region where the ion densities are large and the electron temperature becomes progressively larger than the ion temperature (*Roble*, 1975). The third simulation, which is the normal method in global IT models, used the Joule heating rate with a non-zero neutral photoelectron heating efficiency. Different PHE values were tested (not shown here) and it was found that a Joule simulation with PHE equal to 0.05 had approximately the same globally averaged T_n as the Complete simulation, as shown by the dashed line in Figure 6.1.

Figure 6.2 shows the comparison of the T_n distribution between the three simulations. The top panels show the global horizontal distribution at 300 km of T_n for the three cases. The Joule simulation (i.e., the middle panel) shows a similar distribution of the neutral temperature as the Complete simulation (i.e., the leftmost panel), however, a difference of about 100 K existed globally at 300 km between the Complete and Joule simulations. Although it was expected that it would be the dayside where the Joule heating rate most deviated from the complete neutral-ion collisional heating rate, there was also about 100 K difference in the nightside F-region. This was due to neutral winds advecting the increased temperature from the dayside to the nightside. The comparison of T_n at 180° longitude (middle) and at 300-km altitude above 50 ° latitude (bottom) were shown in Figure 6.2. Large temperature differences are observed in these cuts as well. By increasing the photoelectron heating efficiency to 0.05, the Joule simulation with 0.05 PHE, as shown in the third column, showed a similar global distribution as the Complete simulation excluding photoelectron heating for

the neutrals.

The compensation for the underestimation of T_n in the Joule simulation by the neutral photoelectron heating efficiency was not a coincidence. The approximation of the neutral-ion energy coupling by the Joule heating rate is based on the assumption that the ion temperature is balanced between the energy exchange term and the frictional heating term with the neutrals. However, the heating by ambient electrons could be a non-trivial heat source where the electron density is high, i.e., on the dayside and the F-region at high latitudes (*St-Maurice and Hanson*, 1982; *Killeen et al.*, 1984). In these regions, T_i could deviate from the energy balance assumption as shown in Equation 6.4 due to the ion-electron energy coupling. In other words, T_i was underestimated in the Joule heating rate as a form of the neutral-ion energy coupling. Therefore, the Joule heating rate turned out to be smaller than the complete neutral-ion collisional heating rate in these regions. Furthermore, the difference between the Joule heating rate and the complete neutral-ion collisional heating rate, most likely originates from photoelectrons. This is because photoelectron heating is one of the major heat sources for the ambient thermal electrons on the dayside (*Nagy and Banks*, 1970; *Rasmussen et al.*, 1986; *Smithtro and Solomon*, 2008). The thermal electrons heated through collisions with photoelectrons subsequently transfer thermal energy to ions, leading to a deviation from the neutral energy balance assumption of T_i. Therefore, the non-zero photoelectron heating efficiency used to calculate T_n applied in the case using the Joule heating rate (i.e., the simplified neutral-ion collisional heating rate), mimicked the indirect heating process from photoelectrons to neutrals (through the ions) as a direct heating process.

Figure 6.3 shows the percentage difference of the neutral temperature between the Complete simulation and the Joule simulation, as expressed in Equation 6.10, at 00 UT on Dec 24 (i.e., the end of the last simulation day) at 140 km, 250 km and 400 km. At 140 km, the difference was within 8%, and the northern polar region

had a higher percentage difference than other regions. At 250 km, the difference increased to approximate 12% in the northern hemisphere, and about 8% in the southern hemisphere. At 400 km, the percentage difference maximized at around 15% in the low-latitude region, which was consistent with *St-Maurice and Hanson* (1982); *Killeen et al.* (1984); *Schunk and Nagy* (2009), who found that the ion-electron energy coupling played a less important role for the ion temperature at high latitudes than it did at low and middle latitudes because the electron density was generally lower at high latitudes. Specifically, T_n was underestimated by about 10-12% in the polar regions, and by about 6% on the nightside in the Joule simulation. It was also found that the percentage difference in T_n was higher around the F-region (400 km) than around the E-region (140 km). This could be caused by two reasons: (a) the E-region electron density was generally lower than the F-region density, thus the E-region ion temperature can be better approximated by a balance through energy coupling with the neutrals than the F-region ion temperature; (b) the neutral atmosphere decreases with altitude, which makes the thermosphere at E-region altitudes more difficult to heat through neutral-ion collisional heating (or Joule heating) (*Deng et al.*, 2011).

In Figure 6.4, the colored contours show the difference of the time rate of change of the neutral temperature due to the neutral-ion energy coupling between the Complete and the Joule simulations at 140 km, 250 km and 400 km. The difference was about two orders smaller at 140 km than at 250 km or 400 km, and it was negative around the polar auroral bands at 140km where the ion temperature was slightly higher than the electron temperature due to the large frictional heating with the neutrals. As shown in Equation 6.17, when T_i becomes higher than T_e, the difference of the heating rate becomes negative, meaning that the Joule heating approximation would cause excess heating in these locations. The vertical profile of the ion-neutral frictional heating will be further discussed below. The model limits the electron temperature so that it can be not less than 90% of the ion temperature for stability purposes. At

250 km, the difference reached a peak around 20°- 45° latitude on the dayside, and decreased towards both polar regions. At 400 km, the difference maximized around the geographic equator and generally decreased with latitude, but was relatively large on the dayside and weak on the nightside. There was also a localized maximum in the auroral zone in the southern hemisphere, which could be due to a large deviation of T_i from the energy balance assumption in the auroral band with high electron densities.

Equation 6.17 shows that the difference between the complete neutral-ion collisional heating and the Joule heating rate is proportional to $\frac{n_i}{T_e^{3/2}}(T_e - T_i)$. This term is contoured by the dotted lines in Figure 6.4. The contour of this proportional term generally agrees with that of difference in the time rate of change of T_n at 250 km and 400 km, which, once again, indicates that the difference between the complete neutral-ion collisional heating rate and the Joule heating rate resulted from the lack of consideration of the ion-electron energy coupling in the Joule heating rate.

One feature to note about the global distribution of T_n is that the percentage difference of T_n in Figure 6.3 was greater in the northern hemisphere than in the southern hemisphere. This was caused by two factors. First, as shown in Figure 6.2, the southern hemisphere was generally warmer than the northern hemisphere in December. A greater temperature denominator led to a smaller percentage difference even assuming a similar T_n difference between the two hemispheres. Second, the difference of the time rate of change in the neutral temperature, as shown in Figure 6.4, was generally greater in the northern hemisphere than in the southern hemisphere at 250 km, which indicates a greater T_n difference in the northern hemisphere.

There were some slight inconsistencies between the colored contour and the line contour at 400 km, which is expected because assumptions have been made in the derivation for Equation 6.17, such as using mean masses for simplification, equality of the mean masses between the ions and neutrals and assuming constant ion mean mass. Further, neutral density variations were assumed to be negligible. The uncertainty of

these assumptions could increase in the F-region where transport processes become important, and the variations in mass density could become larger.

Figure 6.5 shows altitudinal profiles of the three major ion thermal sources and losses: the ion-electron heat transfer $Q_T(I-E)$, the ion-neutral frictional heating rate $Q_F(I-N)$ and the negative ion-neutral heat transfer term $-Q_T(I-N)$ as presented in Equations 6.18-6.20 at three different locations. On the dayside, the ion temperature was approximately balanced by the ion-neutral energy coupling below 150 km altitude (i.e., energy gained by frictional heating was lost by heat transfer). The ion-electron heat transfer rate increased quickly with altitude and the ion temperature became a balance between the ion-electron and ion-neutral heat transfer rates around the F-region. The transition region of the energy balance is at approximate 180 km. This means that the Joule heating rate is a good approximation of the neutral-ion collisional heating rate in the E-region on the dayside, but not in the F-region. In the polar region, the ion energy balance was primarily between the frictional heating and heat transfer to the neutrals, until the electron heat transfer became a dominant source of energy above around 350 km. This shows that the Joule heating rate is a relatively good approximation in the high-latitude region below 350 km. On the nightside, the ion-electron heat transfer was one to two orders of magnitude smaller than the other two terms throughout most of the plotted altitudes. Thus, the Joule heating rate was always a good approximate form of the neutral-ion collisional heating rate on the nightside. The frictional heating rate (i.e., yellow line) had a peak in the E-region. This indicates a large heat source for ions by friction with neutrals in this region, such that T_i could be greater than T_e. This helps to explain the negative ion-electron heating rate around 100 km in the polar region and on the nightside. Note that on the nightside between the E-region and F-region, the ion temperature equation is balanced by other terms, such as thermal conduction, instead of only being balanced by friction and heat transfer.

6.2.2 $F10.7$ Dependence

Considering the underestimation of T_n by the Joule heating rate was mainly caused by the neglect of the indirect heating from electrons to neutrals (through ions) and photoelectron heating was the main way to make up this short fall, it may be expected that the performance of the Joule heating rate (with the 5% PHE) was solar-condition dependent. Figure 6.6 shows the evolution of T_n of the same set of simulations as in Figure 6.1 but with $F10.7 = 70$ sfu (left panel) and $F10.7 = 150$ sfu (right panel). When the solar activity was low (i.e., $F10.7 = 70$ sfu), the global averaged T_n was underestimated by about 50 K ($\sim$7%) compared with the Complete simulation in steady state (i.e., at the end of the three simulation days). A photoelectron heating efficiency of 0.035 for the neutral atmosphere compensated for the indirect heating. When the solar radiance was high (i.e., $F10.7 = 150$ sfu), the Joule simulation underestimated the global averaged T_n by about 140 K ($\sim$14%), which could be compensated by a neutral photoelectron heating efficiency of 0.07. During a medium solar condition with $F10.7 = 110$ sfu, as shown in Figure 6.1, the global averaged T_n in the Joule simulation was approximately 90 K ($\sim$11%) cooler than it was in the Complete simulation, which required a photoelectron heating efficiency of 0.05 for compensation. These simulation results suggest a linear relation possibly existing between $F10.7$ and the performance of the Joule heating rate. An increase of 10 sfu of $F10.7$ caused about 1% underestimation of T_n in a simulation using Joule heating rate with no photoelectron heating. The photoelectron heating efficiency increased by 0.015 when $F10.7$ increased from 70 sfu to 110 sfu, and increased by 0.02 when $F10.7$ was increased by 40 from 110 sfu to 150 sfu. This indicates that the PHE for the neutral atmosphere that was required for compensating T_n in Joule simulations tended to increase faster with $F10.7$ during high solar conditions, and the electron-ion heat transfer becomes more important nonlinearly as solar activity increases.

6.3 Discussion and Conclusion

This chapter has discussed the performance of the Joule heating rate as an approximate form of the neutral-ion collisional heating rate in the neutral energy equation in the Global Ionosphere Thermosphere Model. This approximation was valid where the ion-electron collisional heating was negligible and the ion temperature could be approximated by a balance between energy coupling with the neutrals. It has been shown that the global average thermospheric temperature was underestimated by $\sim$11% in the Joule simulation at solar medium (i.e., $F10.7$ equal to 110 sfu) and quiet geomagnetic conditions. The percentage difference of T_n between the two simulations generally decreased from the dayside to the nightside, and from high to low altitudes. At 400 km, the Joule approximation underestimated the neutral temperature by about 15% on the dayside, by about 10-12% in the polar regions, and by about 6% on the nightside. The discrepancy between the Joule heating rate and the Complete neutral-ion collisional heating rate is mainly due to the neglect of the ion-electron heating in the ion energy equation. However, the ion-electron energy coupling can be a non-trivial thermal source for ions in the dayside F-region and in the higher-altitude polar region.

By increasing the photoelectron heating efficiency of the neutral atmosphere, the underestimation of T_n was compensated for quite adequately. A global ionosphere-thermosphere model that used the Joule heating rate as an approximation of the neutral-ion energy coupling, usually applied a PHE for the neutral atmosphere to match model results with observations. However, there has been few studies quantifying the direct heating from photoelectrons to the neutrals. It was found that there exists a roughly linear relation between the performance of the Joule heating approximation and solar activity. Higher $F10.7$ led to a larger discrepancy in T_n in a simulation using Joule heating rate without the employment of a neutral photoelectron heating efficiency. The compensating neutral PHE increased with solar activity

as well. It appeared that the indirect heating of neutrals by electrons increased more efficiently at a high level of solar activity. Besides solar activity, solar wind condition and particle precipitation at high latitude could possibly affect the performance of the Joule heating rate because the convection pattern and auroral activity could effectively change the dynamics of the ionosphere and thermosphere. Further study is needed to investigate the performance of the Joule heating approximation during geomagnetic disturbances. A global IT model should be careful when using the Joule heating rate as an approximate form of the neutral-ion collisional heating rate. Using a fixed neutral PHE to compensate for the loss of the indirect heating from thermal electrons to neutrals may not be proper because the indirect heating from electrons to neutrals can be $F10.7$ dependent. It should also be noted that compensating for one heating source with another may allow quantities such as orbit-averaged mass density to compare quite well with measurement. In addition, since the main area of the temperature difference was on the dayside, and the photoelectron heating also worked on the dayside, any data-model differences caused by issues using the photoelectron heating instead of the complete equation set would be quite subtle during quiet times, as evidence by the comparisons shown here between the two simulations.

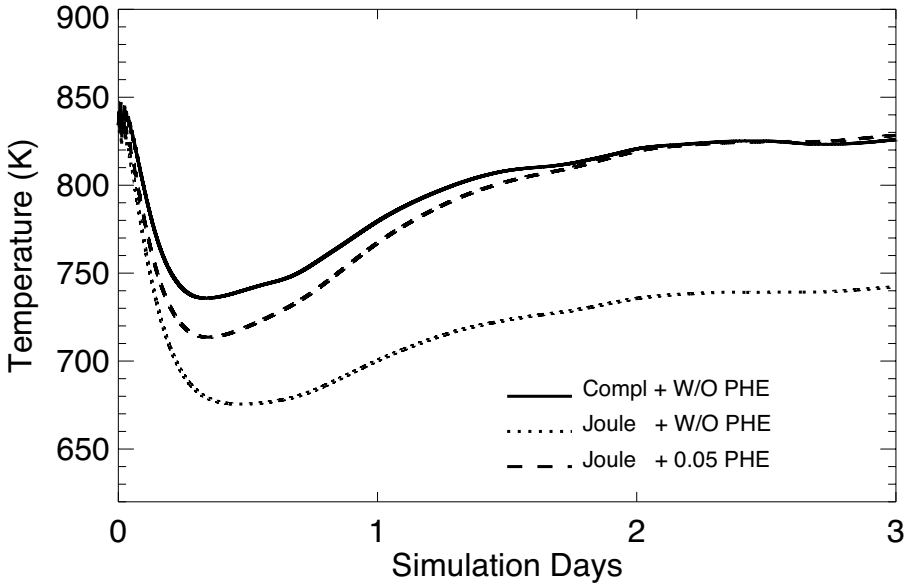

Figure 6.1: The evolution of the global averaged neutral temperature over three simulation days beginning from 00 UT Dec 24, 2012. The solid line shows the simulation using the complete neutral-ion collisional heating rate with PHE equal to zero (termed as the Complete simulation). The dotted line shows the simulation using the Joule heating rate with PHE equal to zero (termed as the Joule simulation). The dashed line shows the simulation using the Joule heating rate with PHE equal to 0.05 (termed with the Joule simulation with 0.05 PHE).

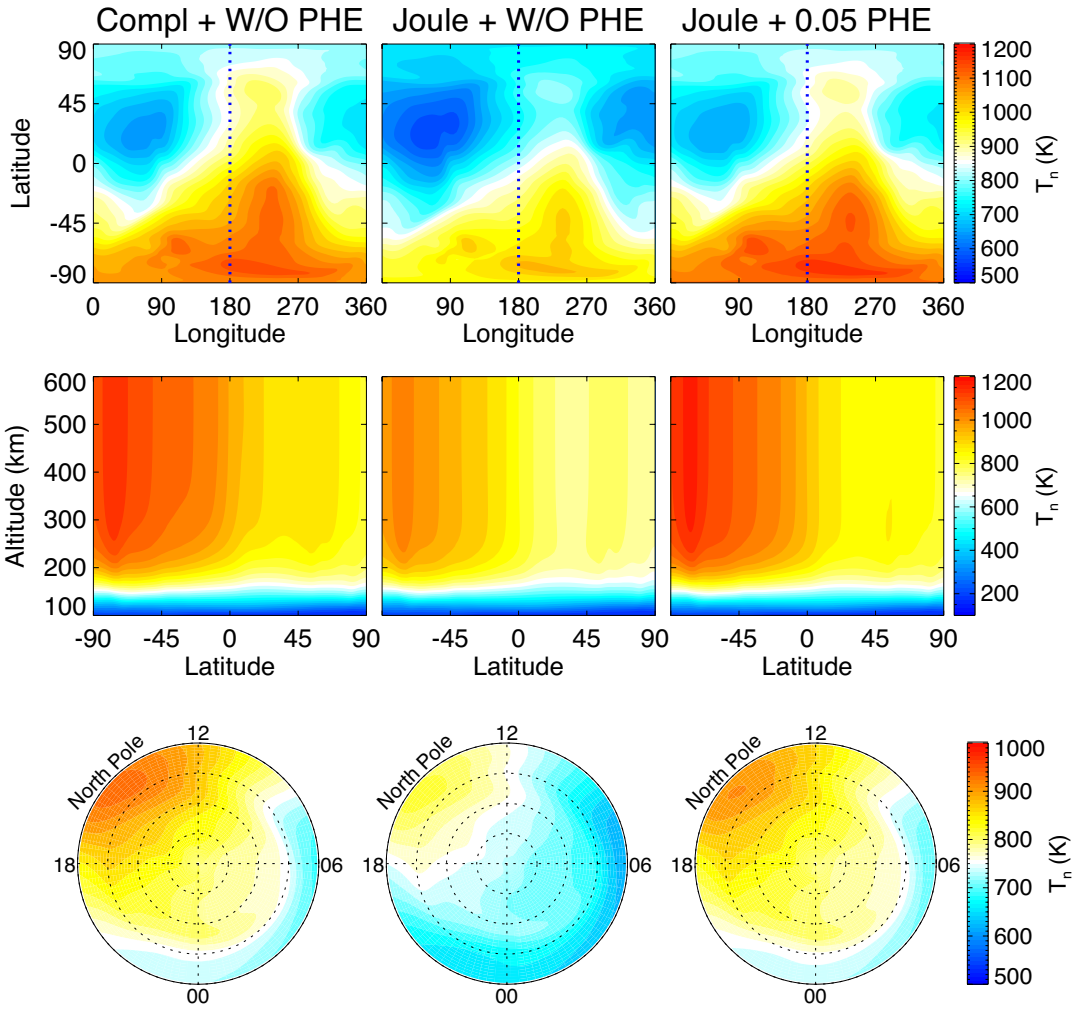

Figure 6.2: The comparison of T_n between the Complete simulation (left column), the Joule simulation (middle column) and the Joule simulation with 0.05 PHE (right column) at 00 UT on Dec 24, 2012. The top row shows the 300-km altitude slice. The second row shows the 180° longitudinal slice. The bottom row shows the north polar cap above 50° latitude.

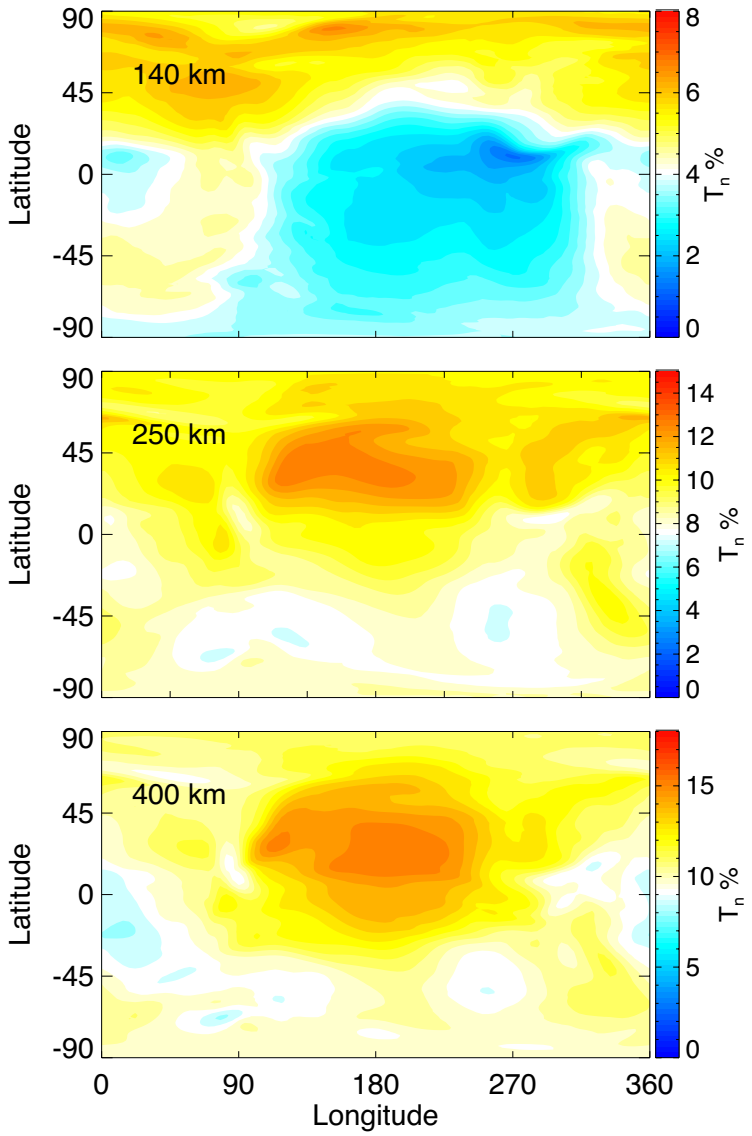

Figure 6.3: The percentage difference of T_n between the Complete and the Joule simulations, i.e., $T_n\% = \frac{(T_n)_C - (T_n)_J}{(T_n)_C} \times 100\%$, at 140 km (top), 250 km (middle) and 400 km (bottom).

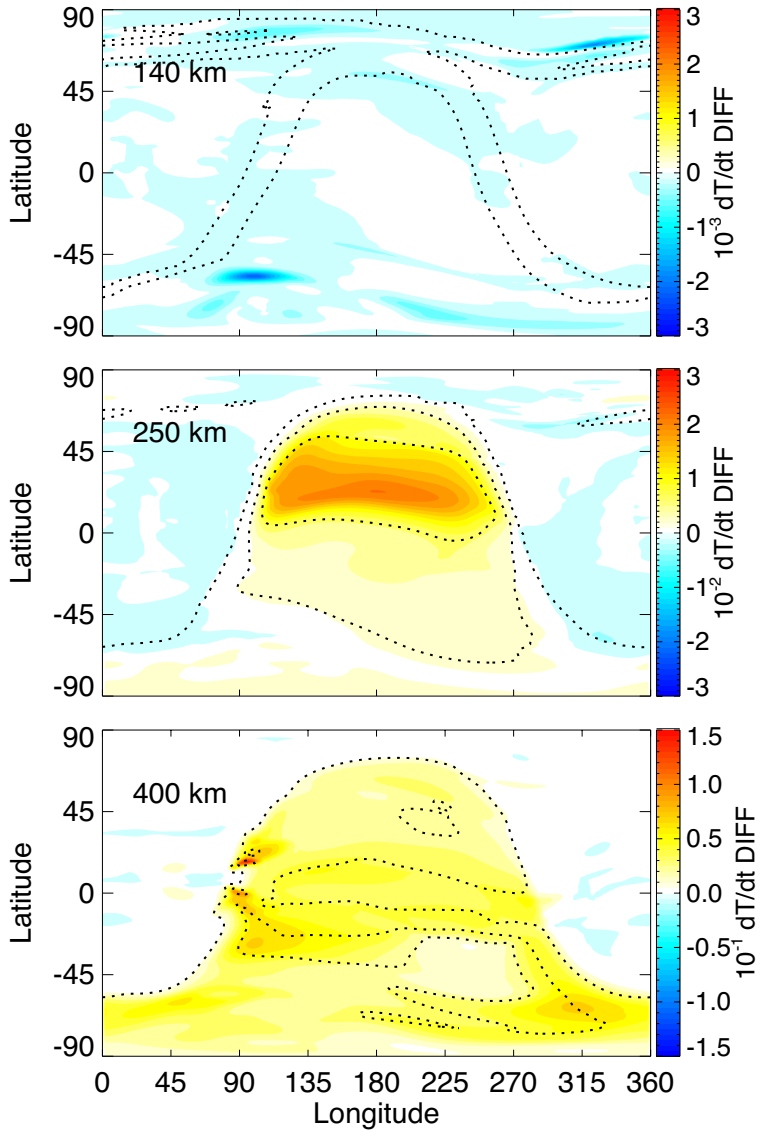

Figure 6.4: The color contour shows the difference of the time rate of change of T_n due to the neutral-ion energy coupling between the Complete and Joule simulations at 140 km (top), 250 km (middle) and 400 km (bottom). The unit is in K$\cdot$ m$^{-3}\cdot$s^{-1}. The dotted line contours the term on the right side of Equation 6.17.

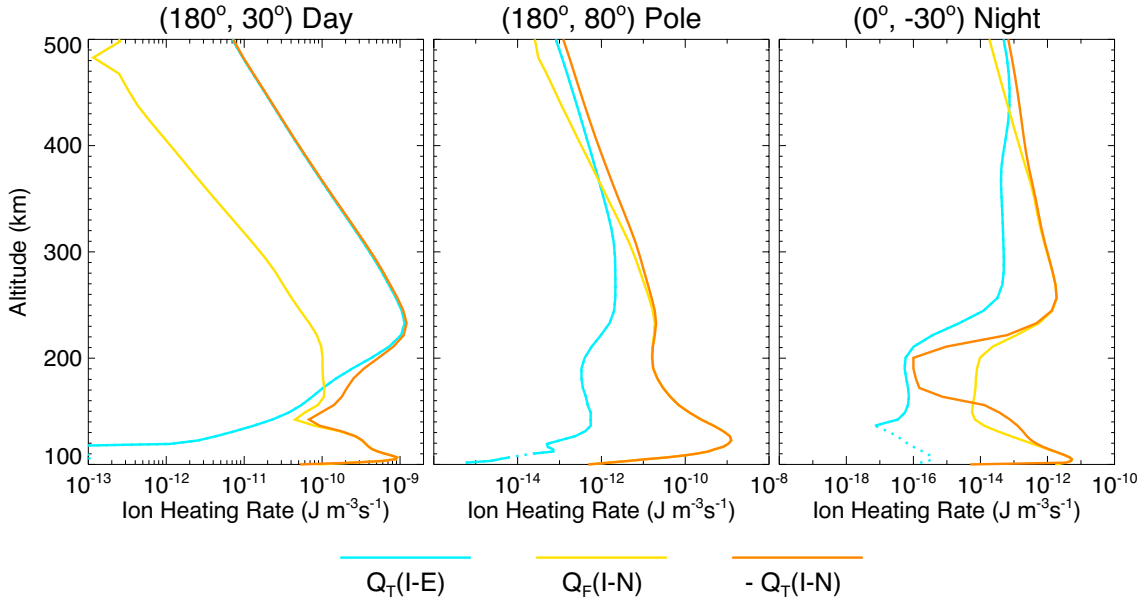

Figure 6.5: The altitudinal profiles of the ion-electron heat transfer rate (blue), the ion-neutral frictional heating rate (yellow) and the negative ion-neutral heat transfer rate (orange) at three geographic locations at 00 UT of the last day of the simulation.

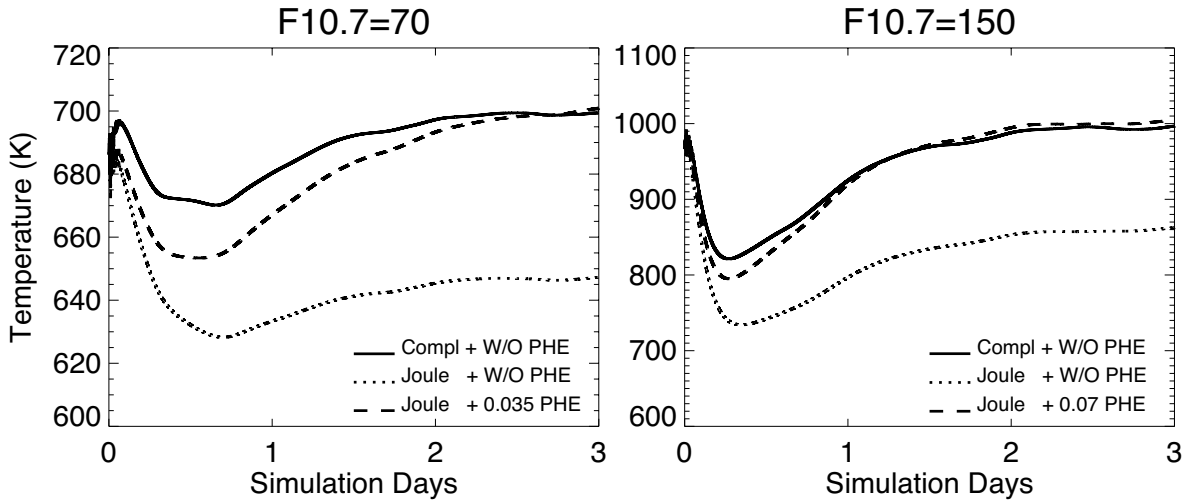

Figure 6.6: In the same format as Figure 6.1. The left and right panels show the temporal evolution of the global averaged T_n when $F10.7 = 70$ sfu and $F10.7 = 150$ sfu respectively.

CHAPTER VII

Conclusion

This dissertation has discussed several important energy channels of the upper atmosphere and how the coupled thermosphere ionosphere system responds these energy inputs by utilizing GITM. Specifically, four topics have been addressed. The first and second studies investigate how the ionospheric density and neutral dynamics respond to sudden energy input by a solar flare. The third study presents an improved electron and ion energy model and the response of the ionospheric temperatures to energy input by solar wind during an idealized substorm. Finally, by using the new model which has neutral, ion and electron energy self-consistently coupled, the Joule heating performance was evaluated.

7.1 TEC response to solar flare

In this study, two of the most important parameters have been found to determine the ionospheric density response to a solar flare, the incident flare energy and the background conditions including geomagnetic field topology and season. First, in order to explore the effect of flare energy, two X5 solar flare events were simulated by grafting the spectra of one flare to the time period of another. In addition, two controlled experiments on simulating the same flare under different background conditions were conducted: one with an ideal dipole field and with the International

Geomagnetic Reference Field (IGRF), and one with equinox and solstice.

The following conclusions were obtained in this study: (1) The total amount of incident energy of the flare dominated the perturbation magnitudes, especially on the dayside; (2) The nightside TEC perturbation started about 12 hours later than other regions. This was because the TEC perturbation did not propagate and remained in the same longitudinal sector until recombination caused it to decay. The TEC perturbation lasted longer than 12 hours but less than 24 hours; (3) Different configurations of the magnetic fields did not significantly change the perturbation except for the nightside. (4) Seasonal effect was important everywhere except for the subsolar region. The polar perturbation is mainly determined by the tilt of the Earth, which controlled the amount of solar irradiance entering the atmosphere in the given hemisphere. The nightside perturbation appeared to be strongly affected by neutral winds.

7.2 Thermospheric waves

Since GITM does not assume hydrostatic equilibrium, it is capable to capture non-hydrostatic waves that occur during intense energy input into the upper atmosphere. The sudden energy input during a solar flare heated up the dayside thermosphere, leading to non-hydro equilibrium locally and launching broad spectra of gravity and acoustic waves. Gravity waves propagated from the dayside to the nightside, and converged at the longitudinal sector which was the anti-subsolar when the flare occurred. The decrease of the "cutoff" region also, from another aspect, indicates that the flare launched significant amount of horizontally propagating gravity waves.

Acoustic waves propagated vertically from the low-altitude source region and increased amplitudes with height in order to conserve the wave energy when the density decreased exponentially. At the time scale of acoustic waves, ion drag was a competitive force to gradient in pressure. Therefore, acoustic waves were significantly

modulated by the ion-neutral momentum coupling term. Specifically, the ion drag caused a shift in the relative phase between the pressure gradient and vertical velocity.

An inverse correlation exists in acoustic waves between the subsolar and anti-subsolar regions, which appeared to be driven by rapid variations of vertical ion flow due to the penetration electric fields. The penetration of solar wind driven fields resulted in an upward and a downward ion drift on the dayside and nightside respectively. The ion flows further forced neutrals to move in the same direction while neutrals tend to restore with the buoyancy frequency.

7.3 Ionospheric temperatures

In this study, the improved ionospheric energy model in GITM was presented. The new model has four main advantages: (1) it is capable of simulating the electron temperature stably using the same time step employed by the neutral atmosphere by using semi-implicit method; (2) it incorporates the thermoelectric heating in a global thermosphere ionosphere model for the first time; (3) it includes the horizontal component of the thermal conduction along field lines for electrons and ions; (4) it enables self-consistent energy coupling between electrons, ions and neutrals.

Comparisons with empirical model and observation data showed improvements of the model on reproducing the magnitudes and basic features of electron and ion temperatures. The new model was used to explore the ionospheric response to an idealized substorm. It was found that T_e and T_i mainly respond during different phases of a substorm. During the expansion phase when the high-latitude convection started increasing, both the E- and F-region ions were heated due to the increase in the ion-neutral frictional heating, while the ion density and electron temperature almost remained unchanged. After the substorm onset, the E-region density immediately responded to the variation in hemispheric power, while the response of the F-region density was delayed by about 30 minutes due to significantly less amount

of particle deposition here. The electron temperature in both the E- and F-region increased immediately due to heating by auroral ionization, which implies that the time scale of electron energy balance is small and the electron thermal conduction efficiently transports heat along the field lines. The magnetospheric heat flux was found to increase electron temperature remarkably in the dusk-to-midnight sector when electron density was relatively low. In addition, the increase in electron temperature tend to reduce the ion loss rate, leading to an enhancement of the F-region density.

7.4 Neutral-Ion energy coupling

The Joule heating rate has been used by some coupled thermosphere ionosphere models as the representation of neutral-ion energy coupling. The approximation is valid where the ion-electron heat transfer is small and the ion temperature can be estimated as a balance by energy coupling with neutrals. This study investigated the performance of this approximation and how this term is internally related to the photoelectron heating efficiency applied to the thermosphere.

Using the Joule heating, the model underestimated the global neutral temperature by about 10% especially in the dayside F-region, at solar medium and under quiet geomagnetic condition. The discrepancy between the Joule heating and the complete neutral-ion collisional heating mainly resulted from the neglect of ion-electron heating in the ion energy equation. By employing a certain value of photoelectron heating efficiency for the neutrals, the underestimation was compensated for quite adequately. The neutral photoelectron heating efficiency employed by a global IT model using Joule heating rate, was basically a compensation for the loss of indirect heating from the thermal electrons to neutrals through ions, and a fixed neutral photoelectron heating efficiency should be carefully selected since it could be solar activity dependent.

7.5 Future work

The studies in this dissertation could be improved in many ways. Here are some examples.

1. In addition to the thermospheric waves, the effect of a solar flare on the neutral wind dynamo could also be explored. It would be expected that the global dynamo electric field could be affected mainly by two ways: (a) the enhancement solar EUV emissions increase ion density on the dayside, thus changing the parallel, Pedersen and Hall conductivities in the ionosphere; (b) the change in the neutral wind due to the flare can cause divergence in ion flow through neutral-ion drag, thus producing currents. Two simulations could be conducted, both with flare spectra but one with dynamo turned on and the other with dynamo off. It would be very interesting to see if the flare-induced dynamo field could cause additional waves, what temporal and spatial scales the waves have and how they propagate and dissipate. In addition, the effect of flare induced dynamo on the ion dynamics could be explored, in order to understand how the dynamo field changes the ion flow and ion density.

2. Electron temperature over the polar cap is poorly simulated in the current model, which could be due to unrealistic representation of heat flux, or any uncertainty in the empirical auroral oval and convection models. Improvement could be achieved by using more realistic auroral (such as OVATION-SM) or convection model. A magnetospheric model could also be coupled to GITM to obtain a better heat flux over the polar cap.

3. Electron temperature could be affected by field-aligned currents (FAC). Since it includes the thermoelectric heating for electrons, the new model is capable of investigating the FAC effect on T_e, and how it depends on solar and magnetic activities. Also, since satellites, such as CHAMP and SWARM, have provided plenty of observations on FAC and ionospheric parameters, model-data comparison becomes available.

4. The Joule heating study could be improved by further investigating how the

IMF-induced electric field affects the thermosphere at high latitudes. Since Joule heating is the dominant way transferring electromagnetic energy from solar wind to thermal energy in the upper atmosphere, understanding the performance of the Joule heating rate as an approximation of neutral-ion energy coupling in global IT models could be particularly important. Interesting questions such as (1) how the Joule heating rate performs in the auroral oval and ion-neutral shear region in the convection zone? (2) how the performance varies with geomagnetic activity? could be addressed.

APPENDICES

APPENDIX A

Source Terms in the Calculation of the Electron Temperature

The electron thermal conductivity λ_e is given by *Schunk and Nagy* (2009):

$$\lambda_e = \frac{7.7 \times 10^5 {T_e}^{5/2}}{1 + 3.22 \times 10^4 \frac{T_e^2}{n_e} \Sigma_n n_n < Q_{en}^{(1)} >}, \tag{A.1}$$

where T_e, n_e and n_n are the electron temperature, electron density and neutral densities, respectively. The densities are in cm^{-3}, and λ_e is in $eV \cdot cm^{-1} s^{-1} K^{-1}$. $< Q_{en}^{(1)} >$ is a Maxwellian average of the momentum transfer cross section, which was simplified as 1.0×10^{-16} cm^2 for all the neutral species in the model. For the theoretical momentum transfer cross sections for some neutral species (not used here), please refer to (*Itikawa*, 1974).

The electron-ion collision frequency is given by *Schunk and Nagy* (2009):

$$\nu_{et} = 54.5 \frac{n_t {Z_t}^2}{{T_e}^{3/2}}, \tag{A.2}$$

where n_t is the number density of ion species t in cm^{-3} and Z_t is the particle charge number.

The electron-neutral collision frequencies are given by (*Itikawa*, 1974; *Schunk and Nagy*, 1980):

$$\nu_{en}(O) = 8.9 \times 10^{-11} n(O)(1 + 5.7 \times 10^{-4} T_e) {T_e}^{1/2}, \tag{A.3}$$

$$\nu_{en}(N_2) = 2.33 \times 10^{-11} n(N_2)(1 - 1.21 \times 10^{-4} T_e) T_e, \tag{A.4}$$

$$\nu_{en}(O_2) = 1.82 \times 10^{-10} n(O_2)(1 + 3.6 \times 10^{-2} {T_e}^{1/2}) {T_e}^{1/2}, \tag{A.5}$$

where $n(O)$, $n(N_2)$ and $n(O_2)$ are the number densities of atomic oxygen, molecular nitrogen and molecular oxygen, respectively. The neutral densities are in cm^{-3}.

The electron cooling rates in $eV \cdot cm^{-3} s^{-1}$ due to inelastic collisions with neutrals are given by the following expressions (densities in cm^{-3}):

N_2 rotation (*Pavlov*, 1998a):

$$L_e(N_2)_{rot} = 3.5 \times 10^{-14} n_e n(N_2)(T_e - T_n) {T_e}^{1/2}. \tag{A.6}$$

where T_n is the neutral temperature.

O_2 rotation (*Pavlov*, 1998b):

$$L_e(O_2)_{rot} = 5.2 \times 10^{-15} n_e n(O_2)(T_e - T_n) {T_e}^{1/2}. \tag{A.7}$$

N_2 vibration (*Pavlov*, 1998a):

$$
\begin{aligned}
L_e(N_2)_{vib} = n_e n(N_2)(1 - e^{(-E_1/T_{vib})}) \\
\sum_{\nu=1}^{10} Q_{0\nu}(1 - e^{(\nu E_1({T_e}^{-1} - T_{vib}^{-1}))}) \\
+ n_e n(N_2)(1 - e^{(-E_1/T_{vib})}) e^{(-E_1/T_{vib})} \\
\sum_{\nu=2}^{9} Q_{1\nu}(1 - e^{((\nu-1)E_1({T_e}^{-1} - {T_{vib}}^{-1}))}),
\end{aligned}
\tag{A.8}
$$

where $E_1 = 3353K$, $T_{vib} = T_n$ and

$$
log_{10}Q_{0\nu} = A_{0\nu} + B_{0\nu}T_e + C_{0\nu}{T_e}^2 + D_{0\nu}{T_e}^3 + F_{0\nu}{T_e}^4 - 16, \tag{A.9}
$$

$$
log_{10}Q_{1\nu} = A_{1\nu} + B_{1\nu}T_e + C_{1\nu}{T_e}^2 + D_{1\nu}{T_e}^3 + F_{1\nu}{T_e}^4 - 16. \tag{A.10}
$$

The A, B, C, D and F coefficients are given in Tables A.1, A.2, and A.3.

Table A.1: Coefficients for calculations of $Q_{0\nu}$ for $300 \leq T_e \leq 1500$ K

ν	$A_{0\nu}$	$B_{0\nu}, K^{-1}$	$C_{0\nu}, K^{-2}$	$D_{0\nu}, K^{-3}$	$F_{0\nu}, K^{-4}$	$\delta_{0\nu}$
1	-6.462	$3.151 \cdot 10^{-2}$	$-4.075 \cdot 10^{-5}$	$2.439 \cdot 10^{-8}$	$-5.479 \cdot 10^{-12}$	0.14

Table A.2: Coefficients for calculations of $Q_{0\nu}$ for $1500 \leq T_e \leq 6000$ K

ν	$A_{0\nu}$	$B_{0\nu}, K^{-1}$	$C_{0\nu}, K^{-2}$	$D_{0\nu}, K^{-3}$	$F_{0\nu}, K^{-4}$	$\delta_{0\nu}$
1	2.025	$8.782 \cdot 10^{-4}$	$2.954 \cdot 10^{-7}$	$-9.562 \cdot 10^{-11}$	$7.252 \cdot 10^{-15}$	0.06
2	-7.066	$1.001 \cdot 10^{-2}$	$-3.066 \cdot 10^{-6}$	$4.436 \cdot 10^{-10}$	$-2.449 \cdot 10^{-14}$	0.08
3	-8.211	$1.092 \cdot 10^{-2}$	$-3.369 \cdot 10^{-6}$	$4.891 \cdot 10^{-10}$	$-2.706 \cdot 10^{-14}$	0.10
4	-9.713	$1.204 \cdot 10^{-2}$	$-3.732 \cdot 10^{-6}$	$5.431 \cdot 10^{-10}$	$-3.008 \cdot 10^{-14}$	0.10
5	-10.353	$1.243 \cdot 10^{-2}$	$-3.850 \cdot 10^{-6}$	$5.600 \cdot 10^{-10}$	$-3.100 \cdot 10^{-14}$	0.13
6	-10.819	$1.244 \cdot 10^{-2}$	$-3.771 \cdot 10^{-6}$	$5.385 \cdot 10^{-10}$	$-2.936 \cdot 10^{-14}$	0.15
7	-10.183	$1.185 \cdot 10^{-2}$	$-3.570 \cdot 10^{-6}$	$5.086 \cdot 10^{-10}$	$-2.769 \cdot 10^{-14}$	0.15
8	-12.698	$1.309 \cdot 10^{-2}$	$-3.952 \cdot 10^{-6}$	$5.636 \cdot 10^{-10}$	$-3.071 \cdot 10^{-14}$	0.15
9	-14.710	$1.409 \cdot 10^{-2}$	$-4.249 \cdot 10^{-6}$	$6.058 \cdot 10^{-10}$	$-3.300 \cdot 10^{-14}$	0.15
10	-17.538	$1.600 \cdot 10^{-2}$	$-4.916 \cdot 10^{-6}$	$7.128 \cdot 10^{-10}$	$-3.941 \cdot 10^{-14}$	0.15

Table A.3: Coefficients for calculations of $Q_{1\nu}$ for $1500 \leq T_e \leq 6000$ K

ν	$A_{1\nu}$	$B_{1\nu}, K^{-1}$	$C_{1\nu}, K^{-2}$	$D_{1\nu}, K^{-3}$	$F_{1\nu}, K^{-4}$	$\delta_{1\nu}$
2	-3.413	$7.326 \cdot 10^{-3}$	$-2.200 \cdot 10^{-6}$	$3.128 \cdot 10^{-10}$	$-1.702 \cdot 10^{-14}$	0.11
3	-4.160	$7.803 \cdot 10^{-3}$	$-2.352 \cdot 10^{-6}$	$3.352 \cdot 10^{-10}$	$-1.828 \cdot 10^{-14}$	0.11
4	-5.193	$8.360 \cdot 10^{-3}$	$-2.526 \cdot 10^{-6}$	$3.606 \cdot 10^{-10}$	$-1.968 \cdot 10^{-14}$	0.12
5	-5.939	$8.807 \cdot 10^{-3}$	$-2.669 \cdot 10^{-6}$	$3.806 \cdot 10^{-10}$	$-2.073 \cdot 10^{-14}$	0.08
6	-8.261	$1.010 \cdot 10^{-2}$	$-3.039 \cdot 10^{-6}$	$4.318 \cdot 10^{-10}$	$-2.347 \cdot 10^{-14}$	0.10
7	-8.185	$1.010 \cdot 10^{-2}$	$-3.039 \cdot 10^{-6}$	$4.318 \cdot 10^{-10}$	$-2.347 \cdot 10^{-14}$	0.12
8	-10.823	$1.199 \cdot 10^{-2}$	$-3.620 \cdot 10^{-6}$	$5.159 \cdot 10^{-10}$	$-2.810 \cdot 10^{-14}$	0.09
9	-11.273	$1.283 \cdot 10^{-2}$	$-3.879 \cdot 10^{-6}$	$5.534 \cdot 10^{-10}$	$-3.016 \cdot 10^{-14}$	0.09

O_2 vibration (*Jones et al.*, 2003):

$$L_e(O_2)_{vib} = n_e n(O_2) Q(T_e)(1 - e^{(2239({T_e}^{-1} - {T_n}^{-1}))}), \tag{A.11}$$

where

$$\begin{aligned} log_{10}Q(T_e) = -19.9171 + 0.0267T_e - 3.9960 \times 10^{-5}{T_e}^2 \\ +3.5187 \times 10^{-8}{T_e}^3 - 1.9228 \times 10^{-11}{T_e}^4 \\ +6.6864 \times 10^{-22}{T_e}^7 - 1.5346 \times 10^{-26}{T_e}^8 \\ +5.0148 \times 10^{-31}{T_e}^9. \end{aligned} \tag{A.12}$$

O fine structure (*Pavlov and Berrington*, 1999):

$$\begin{aligned} L_e(O)_{fine} = n_e n(O) D^{-1}(S_{10}(1 - e^{(98.9({T_e}^{-1} - {T_n}^{-1}))}) \\ +S_{20}(1 - e^{(326.6({T_e}^{-1} - {T_n}^{-1}))}) \\ +S_{21}(1 - e^{(227.7({T_e}^{-1} - 1 - {T_n}^{-1}))}), \end{aligned} \tag{A.13}$$

where

$$D = 5 + e^{(-326.6{T_n}^{-1})} + 3e^{(-227.7{T_n}^{-1})}, \tag{A.14}$$

$$S_{21} = 1.863 \cdot 10^{-11}, \tag{A.15}$$

$$S_{20} = 1.191 \cdot 10^{-11}, \tag{A.16}$$

$$S_{10} = 8.249 \cdot 10^{-16} {T_e}^{0.6} e^{(-227.7{T_n}^{-1})}. \tag{A.17}$$

O(^{1}D) excitation (*Henry et al.*, 1969):

$$L_e(O(^1D)) = 1.57 \times 10^{-12} n_e n(O) e^{(d\frac{T_e - 3000}{3000 T_e})} (e^{(-22713\frac{T_e - T_n}{T_e T_n})} - 1), \tag{A.18}$$

where

$$d = 2.4 \times 10^4 + 0.3(T_e - 1500) - 1.947 \times 10^{-5}(T_e - 1500)(T_e - 4000). \tag{A.19}$$

APPENDIX B

Source Terms in the Calculation of the Ion Temperature

The ion thermal conductivity λ_i is the sum of the thermal conductivity of the major ions weighed by their number densities:

$$\lambda_i = \frac{\sum_t n_t \lambda_t}{\sum_t n_t}, \tag{B.1}$$

where n_t and λ_t are the number density and the thermal conductivity of ion species t. The numerical expression for λ_t is given by (*Schunk and Nagy*, 2009):

$$\lambda_t = 3.1\times10^4 \frac{{T_t}^{5/2}}{{M_t}^{1/2}{Z_t}^4}(1+1.75\sum_{j\neq t}\frac{{Z_j}^2}{{Z_t}^2}\frac{n_j}{n_t}(\frac{M_j}{M_t+M_j})^{1/2}\times\frac{3{M_t}^2+1.6M_tM_j+1.3{M_j}^2}{(M_t+M_j)^2})^{-1}, \tag{B.2}$$

where subscripts t and j refer to different ion species. Note that this expression is obtained for a fully ionized plasma in which ion-ion collisions are dominant. This expression may not apply well to the ionosphere at low altitudes, but is used anyways.

The collision frequencies for resonant ion-neutral interactions are given in Table B (*Banks and Kockarts*, 1973; *Schunk and Nagy*, 2009). The non-resonant collision

frequencies are given by *Schunk and Nagy* (2009):

Table B.1: Collision frequencies for resonant ion-neutral interactions. $T_r = (T_i + T_n)/2$ and densities are in cm^{-3}.

Species	T_r, K	ν_{in}, s^{-1}
N^+, N	> 275	$3.83 \times 10^{-11} n(N) {T_r}^{1/2} (1 - 0.063 log_{10} T_r)^2$
O^+, O	> 235	$3.36 \times 10^{-11} n(O) {T_r}^{1/2} (1 - 0.064 log_{10} T_r)^2$
${N_2}^+, N_2$	> 170	$5.14 \times 10^{-11} n(N_2) {T_r}^{1/2} (1 - 0.069 log_{10} T_r)^2$
${O_2}^+, O_2$	> 800	$2.59 \times 10^{-11} n(N_2) {T_r}^{1/2} (1 - 0.073 log_{10} T_r)^2$

$$\nu(O^+, O_2) = 6.64 \times 10^{10} n(O_2), \tag{B.3}$$

$$\nu(O^+, N_2) = 6.82 \times 10^{10} n(N_2), \tag{B.4}$$

$$\nu({O_2}^+, N_2) = 4.13 \times 10^{10} n(N_2), \tag{B.5}$$

$$\nu({O_2}^+, O) = 2.31 \times 10^{10} n(O), \tag{B.6}$$

where the neutral densities are in cm^{-3}.

APPENDIX C

Charge Exchange and Recombination Reaction Rates

The charge exchange rates κ_1, κ_2, κ_3 are given by (*McFarland et al.*, 1973; *Schunk et al.*, 1975):

$$\kappa_1 = 1.533 \times 10^{-12} - 5.92 \times 10^{-13}(T/300) + 8.6 \times 10^{-14}(T/300)^2 \tag{C.1}$$

for $350K \leq T \leq 1700K$;

$$\kappa_1 = 2.73 \times 10^{-12} - 1.155 \times 10^{-12}(T/300) + 1.483 \times 10^{-13}(T/300)^2 \tag{C.2}$$

for $1700K < T < 6000K$.

$$\begin{aligned}\kappa_2 = 2.82 \times 10^{-11} - 7.74 \times 10^{-12}(T/300) + 1.073 \times 10^{-12}(T/300)^2 \\ -5.17 \times 10^{-14}(T/300)^3 + 9.65 \times 10^{-16}(T/300)^4\end{aligned} \tag{C.3}$$

for $350K \leq T \leq 6000K$.

$$\kappa_3 = 8.36 \times 10^{-13} - 2.02 \times 10^{-13}(T/300) + 6.95 \times 10^{-14}(T/300)^2 \tag{C.4}$$

for $320K \leq T \leq 1500K$;

$$\kappa_3 = 5.33 \times 10^{-13} - 1.64 \times 10^{-14}(T/300) + 4.72 \times 10^{-14}(T/300)^2 - 7.05 \times 10^{-16}(T/300)^3 \tag{C.5}$$

for $1500K < T < 6000K$. In Eqn(C.1)-(C.5), T is the effective temperature, which is given by:

$$T = T(O^+) + \frac{m(O^+)}{m(O^+) + m_r} \frac{m_r - m_b}{3\kappa} u_\perp^2(O^+), \tag{C.6}$$

where

$$m_b = \frac{\sum\limits_m \frac{m_n \nu(O^+, n)}{m(O^+) + m_n}}{\sum\limits_m \frac{\nu(O^+, n)}{m(O^+) + m_n}}, \tag{C.7}$$

m_r is the reactant mass of N_2, O_2 or NO, and $T(O^+)$ is the O^+ temperature. Here, the transport effect was ignored, thus the effective temperature T was reduced to the O^+ temperature (equal the the bulk ion temperature in GITM).

The dissociative recombination rate (in m^{-3}) of NO^+ is given by:

$$\kappa_4 = 4.0 \times 10^{-13} (\frac{300}{Te})^{0.5}. \tag{C.8}$$

The dissociative recombination rate (in m^{-3}) of O_2^+ is given by:

$$\kappa_5 = 2.4 \times 10^{-13} (\frac{300}{Te})^{0.7}. \tag{C.9}$$

The radiative recombination rate (in m^{-3}) of O^+ is given by:

$$\kappa_4 = 3.7 \times 10^{-18}(\frac{300}{Te})^{0.7}. \tag{C.10}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Abdullah, M., H. J. Strangeways, and D. Walsh (2009), Improving ambiguity resolution rate with an accurate ionospheric differential correction, *Journal of Navigation*, *62*(01), 151–166.

Aggarwal, K., N. Nath, and C. Setty (1979), Collision frequency and transport properties of electrons in the ionosphere, *Planetary and Space Science*, *27*(6), 753–768.

Banks, P., and G. Kockarts (1973), Aeronomy, part a, *Academic, San Diego, Calif*, p. 184.

Banks, P., J. Foster, and J. Doupnik (1981), Chatanika radar observations relating to the latitudinal and local time variations of joule heating, *Journal of Geophysical Research: Space Physics (1978–2012)*, *86*(A8), 6869–6878.

Banks, P. M. (1967), Ion temperature in the upper atmosphere, *Journal of Geophysical Research*, *72*(13), 3365–3385.

Barth, C., G. Lu, and R. Roble (2009), Joule heating and nitric oxide in the thermosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *114*(A5).

Baumjohann, W., R. Pelunen, H. Opgenoorth, and E. Nielsen (1981), Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents: Current systems associated with local auroral break-ups, *Planetary and Space Science*, *29*(4), 431–447.

Bekerat, H. A., R. W. Schunk, and L. Scherliess (2007), Estimation of the high-latitude topside electron heat flux using dmsp plasma density measurements, *Journal of atmospheric and solar-terrestrial physics*, *69*(9), 1029–1048.

Bilitza, D. (1986), International reference ionosphere: Recent developments, *Radio Science*, *21*(3), 343–346.

Bilitza, D. (1991), Electron and ion temperature data for ionospheric modelling, *Advances in space Research*, *11*(10), 139–148.

Bilitza, D. (2001), International reference ionosphere 2000, *Radio Science*, *36*(2), 261–275.

Bilitza, D., and W. Hoegy (1990), Solar activity variation of ionospheric plasma temperatures, *Advances in space research*, *10*(8), 81–90.

Bilitza, D., and B. W. Reinisch (2008), International reference ionosphere 2007: improvements and new parameters, *Advances in Space Research*, *42*(4), 599–609.

Bilitza, D., D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. Reinisch (2014), The international reference ionosphere 2012–a model of international collaboration, *Journal of Space Weather and Space Climate*, *4*, A07.

Blewitt, G. (1990), An automatic editing algorithm for gps data, *Geophysical Research Letters*, *17*(3), 199–202.

Brekke, A. (2012), *Physics of the upper polar atmosphere*, Springer Science & Business Media.

Bruinsma, S., D. Tamagnan, and R. Biancale (2004), Atmospheric densities derived from champ/star accelerometer observations, *Planetary and Space Science*, *52*(4), 297–312.

Bruinsma, S., J. M. Forbes, R. S. Nerem, and X. Zhang (2006), Thermosphere density response to the 20–21 november 2003 solar and geomagnetic storm from champ and grace accelerometer data, *Journal of Geophysical Research: Space Physics (1978–2012)*, *111*(A6).

Brund, D. (1927), The period of simple vertical oscillations in the atmosphere, *Quarterly Journal of the Royal Meteorological Society*, *53*(221), 30–32.

Bums, A., T. Killeen, and G. Carignan (1995), Large enhancements in the o/n2 ratio in the evening sector of the winter hemisphere during geomagnetic storms, *Journal of geophysical research*, *100*(A8), 14–661.

Burrell, A. G., A. Goel, A. Ridley, and D. Bernstein (2015), Correction of the photoelectron heating efficiency within the global ionosphere-thermosphere model using retrospective cost model refinement, *Journal of Atmospheric and Solar-Terrestrial Physics*, *124*, 30–38.

Carignan, G., and W. Pinkus (1968), Ogo-f04 experiment description, in *Tech. Note 08041-3-T*, University of Michigan Ann Arbor.

Chamberlin, P. C., T. N. Woods, and F. G. Eparvier (2007), Flare irradiance spectral model (fism): Daily component algorithms and results, *Space Weather*, *5*(7).

Chamberlin, P. C., T. N. Woods, and F. G. Eparvier (2008), Flare irradiance spectral model (fism): Flare component algorithms and results, *Space Weather*, *6*(5).

Chapman, S. (1930), Xxxv. on ozone and atomic oxygen in the upper atmosphere, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *10*(64), 369–383.

Chapman, S. (1931), The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, *Proceedings of the Physical Society*, *43*(1), 26.

Clausen, L., S. Milan, and A. Grocott (2014), Thermospheric density perturbations in response to substorms, *Journal of Geophysical Research: Space Physics*, *119*(6), 4441–4455.

Codrescu, M., T. Fuller-Rowell, and J. Foster (1995), On the importance of e-field variability for joule heating in the high-latitude thermosphere, *Geophysical Research Letters*, *22*(17), 2393–2396.

Cole, K. (1962), Joule heating of the upper atmosphere, *Australian Journal of Physics*, *15*(2), 223–235.

Cole, K. (1971), Electrodynamic heating and movement of the thermosphere, *Planetary and Space Science*, *19*(1), 59–75.

Coster, A., and A. Komjathy (2008), Space weather and the global positioning system, *Space Weather*, *6*(6).

Covington, A. (1948), Solar noise observations on 10.7 centimeters, *Proceedings of the IRE*, *36*(4), 454–457.

Cowley, S., and M. Lockwood (1992), Excitation and decay of solar wind-driven flows in the magnetosphere-ionophere system, in *Annales Geophysicae*, vol. 10, pp. 103–115, Copernicus.

Crowley, G., and R. Meier (2008), Disturbed o/n2 ratios and their transport to middle and low latitudes, *Midlatitude Ionospheric Dynamics and Disturbances*, pp. 221–234.

Da Rosa, A. V. (1966a), The theoretical time-dependent thermal behavior of the ionospheric electron gas, *Journal of Geophysical Research*, *71*(17), 4107–4120.

Da Rosa, A. V. (1966b), The theoretical time-dependent thermal behavior of the ionospheric electron gas, *Journal of Geophysical Research*, *71*(17), 4107–4120.

Dalgarno, A., M. B. McElroy, and R. Moffett (1963), Electron temperatures in the ionosphere, *Planetary and Space Science*, *11*(5), 463–484.

David, M., R. Schunk, and J. Sojka (2011), The effect of downward electron heat flow and electron cooling processes in the high-latitude ionosphere, *Journal of Atmospheric and Solar-Terrestrial Physics*, *73*(16), 2399–2409.

Davies, K. (1990), *Ionospheric radio*, 31, IET.

Deng, Y., and A. J. Ridley (2007), Possible reasons for underestimating joule heating in global models: E field variability, spatial resolution, and vertical velocity, *Journal of Geophysical Research: Space Physics (1978–2012)*, *112*(A9).

Deng, Y., A. D. Richmond, A. J. Ridley, and H.-L. Liu (2008), Assessment of the non-hydrostatic effect on the upper atmosphere using a general circulation model (gcm), *Geophysical Research Letters*, *35*(1).

Deng, Y., T. J. Fuller-Rowell, R. A. Akmaev, and A. J. Ridley (2011), Impact of the altitudinal joule heating distribution on the thermosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *116*(A5).

Dickinson, R. E. (1984), Modeling evapotranspiration for three-dimensional global climate models, *Climate processes and climate sensitivity*, pp. 58–72.

Dickinson, R. E., E. Ridley, and R. Roble (1981), A three-dimensional general circulation model of the thermosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *86*(A3), 1499–1512.

Donnelly, R. (1967), The solar flare radiations responsible for sudden frequency deviations, *Journal of Geophysical Research*, *72*(21), 5247–5256.

Dungey, J. W. (1961), Interplanetary magnetic field and the auroral zones, *Physical Review Letters*, *6*(2), 47.

Emery, B., C. Lathuillere, P. G. Richards, R. Roble, M. Buonsanto, D. Knipp, P. Wilkinson, D. Sipler, and R. Niciejewski (1999), Time dependent thermospheric neutral response tothe 2–11 november 1993 storm period, *Journal of atmospheric and solar-terrestrial physics*, *61*(3), 329–350.

Evans, J. (1969), Theory and practice of ionosphere study by thomson scatter radar, *Proceedings of the IEEE*, *57*(4), 496–530.

Evans, J., and M. Loewenthal (1964), Ionospheric backscatter observations, *Planetary and Space Science*, *12*(10), 913–920.

Evans, J. V. (1965), Ionospheric backscatter observations at millstone hill, *Planetary and Space Science*, *13*(11), 1031–1074.

Farley, D., J. McClure, D. Sterling, and J. Green (1967), Temperature and composition of the equatorial ionosphere, *Journal of Geophysical Research*, *72*(23), 5837–5851.

Fejer, B. G., C. Gonzales, D. Farley, M. Kelley, and R. Woodman (1979), Equatorial electricfields during magnetically disturbed conditions, 1. the effect of the interplanetary magneticfield, *Journal of Geophysical Research*, *84*(A10), 5797–5802.

Forbes, J., F. Marcos, and F. Kamalabadi (1995), Wave structures in lower thermosphere density from satellite electrostatic triaxial accelerometer measurements, *Journal of Geophysical Research: Space Physics (1978–2012)*, *100*(A8), 14,693–14,701.

Forbes, J. M. (2007), Dynamics of the thermosphere, *Journal of the Meteorological Society of Japan*, *85*, 193–213.

Frahm, R., J. Winningham, J. Sharber, R. Link, G. Crowley, E. Gaines, D. Chenette, B. Anderson, and T. Potemra (1997), The diffuse aurora: A significant source of ionization in the middle atmosphere, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *102*(D23), 28,203–28,214.

Frey, H., S. Mende, V. Angelopoulos, and E. Donovan (2004), Substorm onset observations by image-fuv, *Journal of Geophysical Research: Space Physics (1978–2012)*, *109*(A10).

Friedman, J., and F. Herrero (1982), Fabry-perot interferometer measurements of thermospheric neutral wind gradients and reversals at arecibo, *Geophysical Research Letters*, *9*(7), 785–788.

Fritts, D. C., and M. J. Alexander (2003), Gravity wave dynamics and effects in the middle atmosphere, *Reviews of Geophysics*, *41*(1).

Fritts, D. C., and T. S. Lund (2011), Gravity wave influences in the thermosphere and ionosphere: Observations and recent modeling, in *Aeronomy of the Earth's Atmosphere and Ionosphere*, pp. 109–130, Springer.

Fukao, S., W. Oliver, Y. Onishi, T. Takami, T. Sato, T. Tsuda, M. Yamamoto, and S. Kato (1991), F-region seasonal behavior as measured by the mu radar, *Journal of atmospheric and terrestrial physics*, *53*(6), 599–618.

Fuller-Rowell, T., and D. Evans (1987), Height-integrated pedersen and hall conductivity patterns inferred from the tiros-noaa satellite data, *Journal of Geophysical Research: Space Physics (1978–2012)*, *92*(A7), 7606–7618.

Fuller-Rowell, T., and D. Rees (1983), Derivation of a conservation equation for mean molecular weight for a two-constituent gas within a three-dimensional, time-dependent model of the thermosphere, *Planetary and space science*, *31*(10), 1209–1222.

Fuller-Rowell, T., D. Rees, S. Quegan, R. Moffett, and G. Bailey (1988), Simulations of the seasonal and universal time variations of the high-latitude thermosphere and ionosphere using a coupled, three-dimensional, model, in *Ionospheric Modelling*, pp. 189–217, Springer.

Fuller-Rowell, T., M. Codrescu, H. Rishbeth, R. Moffett, and S. Quegan (1996), On the seasonal response of the thermosphere and ionosphere to geomagnetic storms, *Journal of Geophysical Research: Space Physics (1978–2012)*, *101*(A2), 2343–2353.

Fuller-Rowell, T. J., and D. Rees (1980), A three-dimensional time-dependent global model of the thermosphere, *Journal of the Atmospheric Sciences*, *37*(11), 2545–2567.

Gardiner, G. (1969), Origin of the term ionosphere, *Nature*, *224*, 1096.

Garriott, O. K., A. V. da Rosa, M. J. Davis, and O. Villard (1967), Solar flare effects in the ionosphere, *Journal of Geophysical Research*, *72*(23), 6099–6103.

Gary, J., R. Heelis, and J. Thayer (1995), Summary of field-aligned poynting flux observations from de 2, *Geophysical research letters*, *22*(14), 1861–1864.

Gombosi, T. I. (1998), *Physics of the space environment*, Cambridge University Press.

Gordon, W. E. (1958a), Incoherent scattering of radio waves by free electrons with applications to space exploration by radar, *Proceedings of the IRE*, *46*(11), 1824–1829.

Gordon, W. E. (1958b), Incoherent scattering of radio waves by free electrons with applications to space exploration by radar, *Proceedings of the IRE*, *46*(11), 1824–1829.

Gossard, E. E., and W. H. Hooke (1975), Waves in the atmosphere: Atmospheric infrasound and gravity waves-their generation and propagation, *Atmospheric Science*, *2*.

Gross, S. H., C. A. Reber, and F. T. Huang (1984), Large-scale waves in the thermosphere observed by the ae-c satellite, *Geoscience and Remote Sensing, IEEE Transactions on*, (4), 340–352.

Harris, I., and W. Priester (1962), Time-dependent structure of the upper atmosphere, *Journal of the Atmospheric Sciences*, *19*(4), 286–301.

Harris, M., N. Arnold, and A. Aylward (2002), A study into the effect of the diurnal tide on the structure of the background mesosphere and thermosphere using the new coupled middle atmosphere and thermosphere (cmat) general circulation model, in *Annales Geophysicae*, vol. 20, pp. 225–235.

Hays, P., R. Jones, and M. Rees (1973), Auroral heating and the composition of the neutral atmosphere, *Planetary and Space Science*, *21*(4), 559–573.

Hedin, A. (1983), A revised thermospheric model based on mass spectrometer and incoherent scatter data: Msis-83, *Journal of Geophysical Research: Space Physics (1978–2012)*, *88*(A12), 10,170–10,188.

Hedin, A. (1984), Correlations between thermospheric density and temperature, solar euv flux, and 10.7-cm flux variations, *Journal of Geophysical Research: Space Physics (1978–2012)*, *89*(A11), 9828–9834.

Hedin, A., and H. Mayr (1987), Solar euv induced variations in the thermosphere, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *92*(D1), 869–875.

Hedin, A., H. Mayr, C. Reber, N. Spencer, and G. Carignan (1974), Empirical model of global thermospheric temperature and composition based on data from the ogo 6 quadrupole mass spectrometer, *Journal of Geophysical Research*, *79*(1), 215–225.

Hedin, A., et al. (1977), A global thermospheric model based on mass spectrometer and incoherent scatter data msis, 1. n2 density and temperature, *Journal of Geophysical Research*, *82*(16), 2139–2147.

Hedin, A. E. (1987), Msis-86 thermospheric model, *Journal of Geophysical Research: Space Physics (1978–2012)*, *92*(A5), 4649–4662.

Hedin, A. E. (1991), Extension of the msis thermosphere model into the middle and lower atmosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *96*(A2), 1159–1172.

Heelis, R., and W. Coley (1988), Global and local joule heating effects seen by de 2, *Journal of Geophysical Research: Space Physics (1978–2012)*, *93*(A7), 7551–7557.

Heelis, R., J. K. Lowell, and R. W. Spiro (1982), A model of the high-latitude ionospheric convection pattern, *Journal of Geophysical Research: Space Physics (1978–2012)*, *87*(A8), 6339–6345.

Henry, R. J., P. Burke, and A.-L. Sinfailam (1969), Scattering of electrons by c, n, o, n+, o+, and o++, *Physical Review*, *178*(1), 218.

Hernández, G. (1988), *Fabry-Perot Interferometers*, vol. 3, Cambridge University Press.

Hines, C. (1960), Internal atmospheric gravity waves at ionospheric heights, *Canadian Journal of Physics*, *38*(11), 1441–1481.

Hinteregger, H. E., K. Fukui, and B. R. Gilson (1981), Observational, reference and model data on solar euv, from measurements on ae-e, *Geophysical Research Letters*, *8*(11), 1147–1150.

Hodges, R. (1969), Eddy diffusion coefficients due to instabilities in internal gravity waves, *Journal of Geophysical Research*, *74*(16), 4087–4090.

Huang, C.-S., M. Kelley, and D. Hysell (1993), Nonlinear rayleigh-taylor instabilities, atmospheric gravity waves and equatorial spread f, *Journal of Geophysical Research: Space Physics (1978–2012)*, *98*(A9), 15,631–15,642.

Huang, C.-S., J. C. Foster, and M. C. Kelley (2005), Long-duration penetration of the interplanetary electric field to the low-latitude ionosphere during the main phase of magnetic storms, *Journal of Geophysical Research: Space Physics (1978–2012)*, *110*(A11).

Huba, J., G. Joyce, and J. Fedder (2000), Sami2 is another model of the ionosphere (sami2): A new low-latitude ionosphere model, *J. geophys. Res*, *105*(23), 035–23.

Huba, J. D., R. W. Schunk, and G. Khazanov (), *Modeling the Ionosphere-Thermosphere*.

Huffman, R. E. (1969), Absorption cross-sections of atmospheric gases for use in aeronomy, *Canadian Journal of Chemistry*, *47*(10), 1823–1834.

Immel, T. J., G. Crowley, J. D. Craven, and R. G. Roble (2001), Dayside enhancements of thermospheric o/n2 following magnetic storm onset, *Journal of Geophysical Research: Space Physics (1978–2012)*, *106*(A8), 15,471–15,488.

Inhester, B., W. Baumjohann, R. Greenwald, and E. Nielsen (1981), Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. iii- auroral zone currents during the passage of a westward travelling surge, *Journal of Geophysics-Zeitschrift fuer Geophysik*, *49*(3), 155–162.

Itikawa, Y. (1974), Momentum-transfer cross sections for electron collisions with atoms and molecules, *Atomic data and nuclear data tables*, *14*(1), 1–10.

Jacchia, L. (1978), Cira 1972, recent atmospheric models, and improvements in progress, in *COSPAR, 21st Plenary Meeting*, vol. 1.

Jacchia, L., and J. Slowey (1973), A study of the variations in the thermosphere related to solar activity., in *Space Research Conference*, vol. 1, pp. 343–348.

Jacchia, L. G. (1959), Two atmospheric effects in the orbital acceleration of artificial satellites.

Jacka, F. (1984), Application of fabry-perot spectrometers for measurement of upper atmosphere temperatures and winds, in *Handbook for MAP*, vol. 13, pp. 19–40.

Johnson, C. Y. (1966), Ionospheric composition and density from 90 to 1200 kilometers at solar mininum, *jgr*, *71*, 330, doi:10.1029/JZ071i001p00330.

Jones, D. B., L. Campbell, M. J. Bottema, and M. J. Brunger (2003), New electron-energy transfer rates for vibrational excitation of o2, *New Journal of Physics*, *5*(1), 114.

Jones, T. (1971), Vlf phase anomalies due to a solar x-ray flare, *Journal of Atmospheric and Terrestrial Physics*, *33*(6), 963–965.

Kan, J., and L. Lee (1979), Energy coupling function and solar wind-magnetosphere dynamo, *Geophysical Research Letters*, *6*(7), 577–580.

Kato, S. (1956), Horizontal wind systems in the ionospheric e region deduced from the dynamo theory of the geomagnetic s q variation part ii. rotating earth, *Journal of geomagnetism and geoelectricity*, *8*(1), 24–37.

KATO, S. (1957), Horizontal wind systems in the ionospheric e region deduced from the dynamo theory of the geomagnetic s q variation part iv, *Journal of geomagnetism and geoelectricity*, *9*(2), 107–115.

Kelley, M., B. G. Fejer, and C. Gonzales (1979), An explanation for anomalous equatorial ionospheric electric fields associated with a northward turning of the interplanetary magneticfield, *Geophysical Research Letters*, *6*(4), 301–304.

Kelley, M. C. (2009), *The Earth's Ionosphere: Plasma Physics & Electrodynamics*, vol. 96, Academic press.

Kherani, E. A., M. Abdu, E. De Paula, D. Fritts, J. Sobral, and F. de Meneses Jr (2009), The impact of gravity waves rising from convection in the lower atmosphere on the generation and nonlinear evolution of equatorial bubble, *Ann. Geophys*, *27*, 1657–1668.

Kikuchi, T., H. Lühr, K. Schlegel, H. Tachihara, M. Shinohara, and T.-I. Kitamura (2000), Penetration of auroral electric fields to the equator during a substorm, *Journal of Geophysical Research: Space Physics (1978–2012)*, *105*(A10), 23,251–23,261.

Killeen, T., P. Hays, G. Carignan, R. Heelis, W. Hanson, N. Spencer, and L. Brace (1984), Ion-neutral coupling in the high-latitude f region: Evaluation of ion heating terms from dynamics explorer 2, *Journal of Geophysical Research: Space Physics (1978–2012)*, *89*(A9), 7495–7508.

Killeen, T., A. Burns, I. Azeem, S. Cochran, and R. Roble (1997), A theoretical analysis of the energy budget in the lower thermosphere, *Journal of Atmospheric and Solar-Terrestrial Physics*, *59*(6), 675–689.

Klostermeyer, J. (1972), Influence of viscosity, thermal conduction, and ion drag on the propagation of atmospheric gravity waves in the thermosphere, *Zeitschr. Geophys.*, *38*, 881–890.

Kundu, P. K., and I. M. Cohen (2008), *Fluid Mechanics*, 4th ed., Academic Press, United States of America.

Lei, J., R. G. Roble, W. Wang, B. A. Emery, and S.-R. Zhang (2007), Electron temperature climatology at millstone hill and arecibo, *Journal of Geophysical Research: Space Physics (1978–2012)*, *112*(A2).

Leonovich, L., E. Afraimovich, E. Romanova, and A. Taschilin (2002), Estimating the contribution from different ionospheric regions to the tec response to the solar flares using data from the international gps network, *arXiv preprint physics/0202072*.

Leonovich, L., A. Tashchilin, and O. Y. Portnyagina (2010), Dependence of the ionospheric response on the solar flare parameters based on the theoretical modeling and gps data, *Geomagnetism and Aeronomy*, *50*(2), 201–210.

Liemohn, M., J. Kozyra, P. Richards, G. Khazanov, M. Buonsanto, and V. Jordanova (2000), Ring current heating of the thermal electrons at solar maximum, *Journal of Geophysical Research: Space Physics (1978–2012)*, *105*(A12), 27,767–27,776.

Lindzen, R. S. (1967), Thermally driven diurnal tide in the atmosphere, *Quarterly Journal of the Royal Meteorological Society*, *93*(395), 18–42.

Lindzen, R. S. (1968), The application of classical atmospheric tidal theory, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 303, pp. 299–316, The Royal Society.

Lindzen, R. S. (1981), Turbulence and stress owing to gravity wave and tidal breakdown, *Journal of Geophysical Research*, *86*(C10), 9707–9714.

Liu, H., and H. Lühr (2005), Strong disturbance of the upper thermospheric density due to magnetic storms: Champ observations, *Journal of Geophysical Research: Space Physics (1978–2012)*, *110*(A9).

Liu, H., H. Lühr, S. Watanabe, W. Köhler, and C. Manoj (2007a), Contrasting behavior of the thermosphere and ionosphere in response to the 28 october 2003 solar flare, *Journal of Geophysical Research: Space Physics (1978–2012)*, *112*(A7).

Liu, L., H. Le, W. Wan, M. P. Sulzer, J. Lei, and M.-L. Zhang (2007b), An analysis of the scale heights in the lower topside ionosphere based on the arecibo incoherent scatter radar measurements, *Journal of Geophysical Research: Space Physics (1978–2012)*, *112*(A6).

Liu, X., and A. Ridley (2015), A simulation study of the thermosphere mass density response to substorms using gitm, *Journal of Geophysical Research: Space Physics*, *120*(9), 7987–8001.

Liuzzo, L., A. Ridley, N. Perlongo, E. Mitchell, M. Conde, D. Hampton, W. Bristow, and M. Nicolls (2015), High-latitude ionospheric drivers and their effects on wind patterns in the thermosphere, *Journal of Geophysical Research: Space Physics*, *120*(1), 715–735.

Madrigal (2014), The open madrigal initiative, available at `http://www.openmadrigal.org`.

Maeda, K., and S. Kato (1966), Electrodynamics of the ionosphere, *Space Science Reviews*, *5*(1), 57–79.

Maehlum, B., T. Hansen, A. Brekke, O. Holt, and K. Folkestad (1984), Preliminary results from a study of the f-region heating during an intense aurora observed by eiscat, *Journal of Atmospheric and Terrestrial Physics*, *46*(6), 619–623.

Mahajan, K. (1977), Models of electron temperature in the ionospheric f-region using electron density height profiles, *Journal of Atmospheric and Terrestrial Physics*, *39*(5), 637–639.

Mannucci, A. J., B. D. Wilson, and C. D. Edwards (1993), A new method for monitoring the earth's ionospheric total electron content using the gps global network, in *Proceedings of the 6th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 1993)*, pp. 1323–1332.

Marconi, G. (1909), Wireless telegraphic communication, *Nobel Lecture, December*, *11*.

Maruyama, N., A. Richmond, T. Fuller-Rowell, M. Codrescu, S. Sazykin, F. Toffoletto, R. Spiro, and G. Millward (2005), Interaction between direct penetration and disturbance dynamo electric fields in the storm-time equatorial ionosphere, *Geophysical Research Letters*, *32*(17).

Maute, A. (2011), *TIEGCM V 1.94 Model Description*, National Center for Atmospheric Research, Boulder, CO.

Mayr, H., I. Harris, F. Herrero, N. Spencer, F. Varosi, and W. Pesnell (1990), Thermospheric gravity waves: observations and interpretation using the transfer function model (tfm), *Space Science Reviews*, *54*(3-4), 297–375.

McClure, J. (1971), Thermospheric temperature variations inferred from incoherent scatter observations, *Journal of Geophysical Research*, *76*(13), 3106–3115.

McFarland, M., D. Albritton, F. Fehsenfeld, E. Ferguson, and A. Schmeltekopf (1973), Flow-drift technique for ion mobility and ion-molecule reaction rate constant measurements. ii. positive ion reactions of n+, o+, and h2+ with o2 and o+ with n2 from thermal to [inverted lazy s] 2 ev, *The Journal of chemical physics*, *59*(12), 6620–6628.

Meada, S. (1992), Heat budget of the thermosphere and temperature variations during the recovery phase of a geomagnetic storm, *jgr*, *97*, 14,947–14,957, doi:10.1029/92JA01368.

Mendillo, M., et al. (1974), Behavior of the ionospheric f region during the great solar flare of august 7, 1972, *Journal of Geophysical Research*, *79*(4), 665–672.

Millward, G., R. Moffett, S. Quegan, and T. Fuller-Rowell (1996), A coupled thermosphere-ionosphere-plasmasphere model (ctip), *Solar Terrestrial Energy Program (STEP) Handbook of Ionospheric Models*, pp. 239–279.

Nagy, A., and P. Banks (1970), Photoelectron fluxes in the ionosphere, *Journal of Geophysical Research*, *75*(31), 6260–6270.

Nier, A. O., W. E. Potter, D. R. Hickman, and K. Mauersberger (1973), The open-source neutral-mass spectrometer on atmosphere explorer-c, -d, and -e, *Radio Science*, *8*(4), 271–276, doi:10.1029/RS008i004p00271.

Nishida, A. (1968), Coherence of geomagnetic dp2 fluctuations with interplanetary magnetic variations, *J. geophys. Res*, *73*(17), 5549–5559.

Ohtani, S.-i. (2000), *Magnetospheric current systems*, vol. 118, American Geophysical Union.

Oliver, W., Y. Otsuka, M. Sato, T. Takami, and S. Fukao (1997), A climatology of f region gravity wave propagation over the middle and upper atmosphere radar, *Journal of Geophysical Research: Space Physics (1978–2012)*, *102*(A7), 14,499–14,512.

Opgenoorth, H., I. Hägström, P. Williams, and G. Jones (1990), Regions of strongly enhanced perpendicular electric fields adjacent to auroral arcs, *Journal of Atmospheric and Terrestrial Physics*, *52*(6), 449–458.

Otsuka, Y., S. Kawamura, N. Balan, S. Fukao, and G. Bailey (1998), Plasma temperature variations in the ionosphere over the middle and upper atmosphere radar, *Journal of Geophysical Research: Space Physics (1978–2012)*, *103*(A9), 20,705–20,713.

Oyama, K.-I., S. Watanabe, Y. Su, T. Takahashi, and K. Hirao (1996), Season, local time, and longitude variations of electron temperature at the height of 600 km in the low latitude region, *Advances in Space Research*, *18*(6), 269–278.

Pavlov, A. (1998a), New electron energy transfer rates for vibrational excitation of n2, in *Annales Geophysicae*, vol. 16, pp. 176–182, Springer.

Pavlov, A. (1998b), New electron energy transfer and cooling rates by excitation of o2, in *Annales Geophysicae*, vol. 16, pp. 1007–1013, Springer.

Pavlov, A., and K. Berrington (1999), Cooling rate of thermal electrons by electron impact excitation of fine structure levels of atomic oxygen, in *Annales Geophysicae*, vol. 17, pp. 919–924, Springer.

Pawlowski, D. J., and A. J. Ridley (2008), Modeling the thermospheric response to solar flares, *Journal of Geophysical Research: Space Physics (1978–2012)*, *113*(A10).

Pawlowski, D. J., and A. J. Ridley (2009), Modeling the ionospheric response to the 28 october 2003 solar flare due to coupling with the thermosphere, *Radio Science*, *44*(1).

Press, F., and D. Harkrider (1962), Propagation of acoustic-gravity waves in the atmosphere, *Journal of Geophysical Research*, *67*(10), 3889–3908.

Prölss, G. (1980), Magnetic storm associated perturbations of the upper atmosphere: Recent results obtained by satellite-borne gas analyzers, *Reviews of Geophysics*, *18*(1), 183–202.

Qian, L., A. G. Burns, P. C. Chamberlin, and S. C. Solomon (2010), Flare location on the solar disk: Modeling the thermosphere and ionosphere response, *Journal of Geophysical Research: Space Physics (1978–2012)*, *115*(A9).

Qian, L., A. G. Burns, S. C. Solomon, and P. C. Chamberlin (2012), Solar flare impacts on ionospheric electrodyamics, *Geophysical Research Letters*, *39*(6).

Rasmussen, C., R. W. Schunk, J. J. Sojka, V. B. Wickwar, O. Beaujardiere, J. Foster, J. Holt, D. Evans, and E. Nielsen (1986), Comparison of simultaneous chatanika and millstone hill observations with ionospheric model predictions, *Journal of Geophysical Research: Space Physics (1978–2012)*, *91*(A6), 6986–6998.

Rees, D., and T. Fuller-Rowell (1989), The response of the thermosphere and ionosphere to magnetospheric forcing, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *328*(1598), 139–171.

Rees, M. H. (1989), *Physics and chemistry of the upper atmosphere*, vol. 1, Cambridge University Press.

Reiff, P. H. (1982), Sunward convection in both polar caps, *Journal of Geophysical Research: Space Physics (1978–2012)*, *87*(A8), 5976–5980.

Reifman, A., and W. G. Dow (1949), Dynamic probe measurements in the ionosphere, *Physical Review*, *76*(7), 987.

Reigber, C., H. Lühr, and P. Schwintzer (2002), Champ mission status, *Advances in Space Research*, *30*(2), 129–134.

Richard, P., G. Fennelly, and D. Torr (1994), Euvac: a solar euv flux model for aeronomic calculation, *J. Geophys. Res*, *99*, 8981–8992.

Richards, P., and D. Torr (1983), A simple theoretical model for calculating and parameterizing the ionospheric photoelectron flux, *Journal of Geophysical Research: Space Physics (1978–2012)*, *88*(A3), 2155–2162.

Richards, P., J. Fennelly, and D. Torr (1994), Correction to euvac: A solar euv flux model for aeronomic calculations, *Journal of Geophysical Research: Space Physics (1978–2012)*, *99*(A7), 13,283–13,283.

Richmond, A. (1979), Large-amplitude gravity wave energy production and dissipation in the thermosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *84*(A5), 1880–1890.

Richmond, A. (1989), Modeling the ionosphere wind dynamo: A review, in *Quiet Daily Geomagnetic Fields*, pp. 413–435, Springer.

Richmond, A. (1995), Ionospheric electrodynamics using magnetic apex coordinates., *Journal of geomagnetism and geoelectricity*, *47*(2), 191–212.

Richmond, A., and S. Matsushita (1975), Thermospheric response to a magnetic substorm, *Journal of Geophysical Research*, *80*(19), 2839–2850.

Richmond, A., E. Ridley, and R. Roble (1992), A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophysical Research Letters*, *19*(6), 601–604.

Rideout, W., and A. Coster (2006), Automated gps processing for global total electron content data, *GPS Solutions*, *10*(3), 219–228.

Ridley, A. (2000), Estimations of the uncertainty in timing the relationship between magnetospheric and solar wind processes, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(9), 757–771.

Ridley, A., Y. Deng, and G. Toth (2006), The global ionosphere–thermosphere model, *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(8), 839–864.

Rishbeth, H. (1988), Basic physics of the ionosphere:a tutorial review, *Journal of the Institution of Electronic and Radio Engineers*, *58*, S207–S223, doi:10.1049/jiere.1988.0060.

Rishbeth, H., and I. Müller-Wodarg (1999), Vertical circulation and thermospheric composition: a modelling study, in *Annales Geophysicae*, vol. 17, pp. 794–805, Springer.

Rishbeth, H., and C. Setty (1961), The f-layer at sunrise, *Journal of Atmospheric and Terrestrial Physics*, *20*(4), 263–276.

Rishbeth, H., T. Fuller-Rowell, and D. Rees (1987), Diffusive equilibrium and vertical motion in the thermosphere during a severe magnetic storm: a computational study, *Planetary and space science*, *35*(9), 1157–1165.

Rishbeth, H., I. Müller-Wodarg, L. Zou, T. Fuller-Rowell, G. Millward, R. Moffett, D. Idenden, and A. Aylward (2000), Annual and semiannual variations in the ionospheric f2-layer: Ii. physical discussion, in *Annales Geophysicae*, vol. 18, pp. 945–956, Springer.

Roble, R. (1975), The calculated and observed diurnal variation of the ionosphere over millstone hill on 23–24 march 1970, *Planetary and Space Science*, *23*(7), 1017–1033.

Roble, R., and B. Emery (1983), On the global mean temperature of the thermosphere, *Planetary and Space Science*, *31*(6), 597–614.

Roble, R., and E. Ridley (1987), An auroral model for the ncar thermospheric general circulation model (tgcm), in *Annales Geophysicae*, vol. 5, pp. 369–382.

Roble, R., E. Ridley, and R. Dickinson (1987), On the global mean structure of the thermosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *92*(A8), 8745–8758.

Roble, R., E. C. Ridley, A. Richmond, and R. Dickinson (1988), A coupled thermosphere/ionosphere general circulation model, *Geophysical Research Letters*, *15*(12), 1325–1328.

Rodger, A., G. Wells, R. Moffett, and G. Bailey (2001), The variability of joule heating, and its effects on the ionosphere and thermosphere, in *Annales Geophysicae*, vol. 19, pp. 773–781.

Römer, M. (1967), Geomagnetic activity effect derived from explorer 9 drag data, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *262*(1124), 185–194.

Sastri, J. H., J. Rao, D. Rao, and B. Pathan (2001), Daytime equatorial geomagnetic h field response to the growth phase and expansion phase onset of isolated substorms: Case studies and their implications, *Journal of Geophysical Research: Space Physics (1978–2012)*, *106*(A12), 29,925–29,933.

Sayers, J. (1970), In-situ probes for ionospheric investigations, *Journal of Atmospheric and Terrestrial Physics*, *32*(4), 663–691.

Schmahl, E., and M. Kundu (1995), Microwave proxies for sunspot blocking and total irradiance, *Journal of Geophysical Research: Space Physics (1978–2012)*, *100*(A10), 19,851–19,864.

Schunk, R. (1975), Transport equations for aeronomy, *Planetary and Space Science*, *23*(3), 437–485.

Schunk, R. (1988), A mathematical model of the middle and high latitude ionosphere, in *Ionospheric Modelling*, pp. 255–303, Springer.

Schunk, R., and A. Nagy (1980), Ionospheres of the terrestrial planets, *Reviews of Geophysics*, *18*(4), 813–852.

Schunk, R., and A. Nagy (2009), *Ionospheres: Physics, Plasma Physics, and Chemistry*, 2rd ed., Cambridge University Press, United States of America.

Schunk, R., and A. F. Nagy (1978a), Electron temperatures in the f region of the ionosphere: Theory and observations, *Reviews of Geophysics*, *16*(3), 355–399.

Schunk, R., and A. F. Nagy (1978b), Electron temperatures in the f region of the ionosphere: Theory and observations, *Reviews of Geophysics*, *16*(3), 355–399.

Schunk, R., W. Raitt, and P. Banks (1975), Effect of electric fields on the daytime high-latitude e and f regions, *Journal of Geophysical Research*, *80*(22), 3121–3130.

Schunk, R. W., and J. J. Sojka (1982), Ion temperature variations in the daytime high-latitude f region, *Journal of Geophysical Research: Space Physics (1978–2012)*, *87*(A7), 5169–5183.

Schunk, R. W., and J. C. Walker (1970), Transport properties of the ionospheric electron gas, *Planetary and Space Science*, *18*(11), 1535–1550.

Schunk, R. W., J. J. Sojka, and M. Bowline (1986a), Theoretical study of the electron temperature in the high-latitude ionosphere for solar maximum and winter conditions, *Journal of Geophysical Research: Space Physics (1978–2012)*, *91*(A11), 12,041–12,054.

Schunk, R. W., J. J. Sojka, and M. Bowline (1986b), Theoretical study of the electron temperature in the high-latitude ionosphere for solar maximum and winter conditions, *Journal of Geophysical Research: Space Physics (1978–2012)*, *91*(A11), 12,041–12,054.

Schunk, R. W., et al. (2004), Global assimilation of ionospheric measurements (gaim), *Radio Science*, *39*(1).

Smith, F., and C. Smith (1972), Numerical evaluation of chapman's grazing incidence integral ch (x, χ), *Journal of Geophysical Research*, *77*(19), 3592–3597.

Smithtro, C., and S. Solomon (2008), An improved parameterization of thermal electron heating by photoelectrons, with application to an x17 flare, *Journal of Geophysical Research: Space Physics (1978–2012)*, *113*(A8).

Sojka, J. J. (1989), Global scale, physical models of the f region ionospere, *Reviews of Geophysics*, *27*(3), 371–403.

Solomon, S. C., and V. J. Abreu (1989), The 630 nm dayglow, *Journal of Geophysical Research: Space Physics (1978–2012)*, *94*(A6), 6817–6824.

Solomon, S. C., P. B. Hays, and V. J. Abreu (1988), The auroral 6300 å emission: Observations and modeling, *Journal of Geophysical Research: Space Physics (1978–2012)*, *93*(A9), 9867–9882.

Southwood, D. (1977), The role of hot plasma in magnetospheric convection, *Journal of Geophysical Research*, *82*(35), 5512–5520.

Spaceweatherlive (2014), Real-time auroral activity and solar activity, available at `http://www.spaceweatherlive.com/en/solar-activity`.

St-Maurice, J., and R. Schunk (1979), Ion velocity distributions in the high-latitude ionosphere, *Reviews of Geophysics and Space Physics*, *17*(1), 99–133.

St-Maurice, J.-P., and W. Hanson (1982), Ion frictional heating at high latitudes and its possible use for an in situ determination of neutral thermospheric winds and temperatures, *Journal of Geophysical Research: Space Physics (1978–2012)*, *87*(A9), 7580–7602.

Stephenson, F. R., D. M. Willis, and T. J. Hallinan (2004), The earliest datable observation of the aurora borealis, *Astronomy & Geophysics*, *45*(6), 6–15.

Stolarski, R., P. Hays, and R. Roble (1975), Atmospheric heating by solar euv radiation, *Journal of Geophysical Research*, *80*(16), 2266–2276.

Stonehocker, G. (1970), Advanced telecommunication forecasting technique, *AGARD Conference Proceedings*, *49*(27), 1.

Strangeway, R. J. (2012), The equivalence of joule dissipation and frictional heating in the collisional ionosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *117*(A2).

Svenes, K., B. Maehlum, U. Løvhaug, M. Rietveld, and C. Hall (1992a), An investigation of two different electron heating events in the ionospheric f-layer, *Advances in Space Research*, *12*(6), 241–245.

Svenes, K., B. Maehlum, J. Trøim, G. Holmgren, R. Arnoldy, U. Løvhaug, M. Rietveld, and C. Hall (1992b), Combined rocket and ground observations of electron heating in the ionospheric f-layer, *Planetary and space science*, *40*(7), 901–912.

Swartz, W. E., and J. S. Nisbet (1972), Revised calculations of f region ambient electron heating by photoelectrons, *Journal of Geophysical Research*, *77*(31), 6259–6261.

SWPC (2013), Historical swp products from 1996, available at `http://www.swpc.noaa.gov/ftpmenu/`.

SWPC (2015), F10.7 cm radio emissions, available at `www.swpc.noaa.gov/phenomena/f107-cm-radio-emissions`.

Tarpley, J. (1970), The ionospheric wind dynamoii: Solar tides, *Planetary and Space Science*, *18*(7), 1091–1103.

Thayer, J. (1998), Height-resolved joule heating rates in the high-latitude e region and the influence of neutral winds, *Journal of Geophysical Research: Space Physics (1978–2012)*, *103*(A1), 471–487.

Thayer, J., J. Vickrey, R. Heelis, and J. Gary (1995), Interpretation and modeling of the high-latitude electromagnetic energy flux, *Journal of Geophysical Research: Space Physics (1978–2012)*, *100*(A10), 19,715–19,728.

Thayer, J. P., and J. Semeter (2004), The convergence of magnetospheric energy flux in the polar atmosphere, *Journal of atmospheric and solar-terrestrial physics*, *66*(10), 807–824.

Tobiska, W. K., and C. A. Barth (1990), A solar euv flux model, *Journal of Geophysical Research: Space Physics (1978–2012)*, *95*(A6), 8243–8251.

Torr, M., D. Torr, and P. Richards (1980a), The solar ultraviolet heating efficiency of the midlatitude thermosphere, *Geophysical Research Letters*, *7*(5), 373–376.

Torr, M. R., P. Richards, and D. Torr (1980b), A new determination of the ultraviolet heating efficiency of the thermosphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *85*(A12), 6819–6826.

Torr, M. R., D. Torr, P. Richards, and S. Yung (1990), Mid-and low-latitude model of thermospheric emissions: 1. o+ (^{2}p) 7320 å and n2 (2p) 3371 å, *Journal of Geophysical Research: Space Physics (1978–2012)*, *95*(A12), 21,147–21,168.

Truhlik, V., L. Triskova, D. Bilitza, and K. Podolska (2009), Variations of daytime and nighttime electron temperature and heat flux in the upper ionosphere, topside ionosphere and lower plasmasphere for low and high solar activity, *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*(17), 2055–2063.

Tsurutani, B., O. Verkhoglyadova, A. Mannucci, G. Lakhina, G. Li, and G. Zank (2009), A brief review of solar flare effects on the ionosphere, *Radio Science*, *44*(1).

Tsurutani, B., et al. (2005), The october 28, 2003 extreme euv solar flare and resultant extreme ionospheric effects: Comparison to other halloween events and the bastille day event, *Geophysical Research Letters*, *32*(3).

Vadas, S. L., and D. C. Fritts (2002), The importance of spatial variability in the generation of secondary gravity waves from local body forces, *Geophysical research letters*, *29*(20), 45–1.

Väisälä, V. (1932), *Über die Wirkung der Windschwankungen auf die Pilotbeobachtungen*, Meteorologischen Institut der Universität Helsinki.

VanZandt, T., W. Clark, and J. Warnock (1972), Magnetic apex coordinates: A magnetic coordinate system for the ionospheric f 2 layer, *Journal of Geophysical Research*, *77*(13), 2406–2411.

Vasyliunas, V. M. (1972), The interrelationship of magnetospheric processes, in *Earths Magnetospheric Processes*, pp. 29–38, Springer.

Wahlund, J.-E., H. Opgenoorth, F. Forme, M. Persson, I. Häggström, and J. Lilensten (1993), Electron energization in the topside auroral ionosphere: on the importance of ion-acoustic turbulence, *Journal of atmospheric and terrestrial physics*, *55*(4), 623–645.

Wan, W., L. Liu, H. Yuan, B. Ning, and S. Zhang (2005), The gps measured sitec caused by the very intense solar flare on july 14, 2000, *Advances in Space Research*, *36*(12), 2465–2469.

Wang, W., T. L. Killeen, A. G. Burns, and R. G. Roble (1999), A high-resolution, three-dimensional, time dependent, nested grid model of the coupled thermosphere–ionosphere, *Journal of Atmospheric and Solar-Terrestrial Physics*, *61*(5), 385–397.

Wang, W., A. Burns, and T. Killeen (2006), A numerical study of the response of ionospheric electron temperature to geomagnetic activity, *Journal of Geophysical Research: Space Physics (1978–2012)*, *111*(A11).

Wang, W., J. Lei, A. G. Burns, M. Wiltberger, A. D. Richmond, S. C. Solomon, T. L. Killeen, E. R. Talaat, and D. N. Anderson (2008), Ionospheric electric field variations during a geomagnetic storm simulated by a coupled magnetosphere ionosphere thermosphere (cmit) model, *Geophysical Research Letters*, *35*(18).

Weimer, D. (2005), Improved ionospheric electrodynamic models and application to calculating joule heating rates, *Journal of Geophysical Research: Space Physics (1978–2012)*, *110*(A5).

Willmore, A. (1970), Electron and ion temperatures in the ionosphere, *Space Science Reviews*, *11*(5), 607–670.

Woods, T. N., F. G. Eparvier, J. Fontenla, J. Harder, G. Kopp, W. E. McClintock, G. Rottman, B. Smiley, and M. Snow (2004), Solar irradiance variability during the october 2003 solar storm period, *Geophysical research letters*, *31*(10).

Woods, T. N., et al. (1998), Timed solar euv experiment, in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pp. 180–191, International Society for Optics and Photonics.

Worden, J. R., T. N. Woods, and K. W. Bowman (2001), Far-ultraviolet intensities and center-to-limb variations of active regions and quiet sun using uars solstice irradiance measurements and ground-based spectroheliograms, *The Astrophysical Journal*, *560*(2), 1020.

Xiong, B., W. Wan, B. Ning, F. Ding, L. Hu, and Y. Yu (2014), A statistic study of ionospheric solar flare activity indicator, *Space Weather*, *12*(1), 29–40.

Xiong, B., et al. (2011), Ionospheric response to the x-class solar flare on 7 september 2005, *Journal of Geophysical Research: Space Physics (1978–2012)*, *116*(A11).

Yeh, K. C., and C. H. Liu (1974a), Acoustic-gravity waves in the upper atmosphere, *Reviews of Geophysics*, *12*(2), 193–216.

Yeh, K. C., and C. H. Liu (1974b), Acoustic-gravity waves in the upper atmosphere, *Reviews of Geophysics*, *12*(2), 193–216.

Yeoman, T., J. Davies, N. Wade, G. Provan, and S. E. Milan (2000), Combined cutlass, eiscat and esr observations of ionospheric plasma flows at the onset of an isolated substorm, in *Annales Geophysicae*, vol. 18, pp. 1073–1087, Springer.

Zhang, B.-C., Y. Kamide, and R.-Y. Liu (2003), Response of electron temperature to field-aligned current carried by thermal electrons: A model, *Journal of Geophysical Research: Space Physics (1978–2012)*, *108*(A5).

Zhang, D., and Z. Xiao (2003), Study of the ionospheric total electron content response to the great flare on 15 april 2001 using the international gps service network for the whole sunlit hemisphere, *Journal of Geophysical Research: Space Physics (1978–2012)*, *108*(A8).

Zhang, D., and Z. Xiao (2005a), Study of ionospheric response to the 4b flare on 28 october 2003 using international gps service network data, *Journal of Geophysical Research: Space Physics (1978–2012)*, *110*(A3).

Zhang, D., and Z. Xiao (2005b), Study of ionospheric response to the 4b flare on 28 october 2003 using international gps service network data, *Journal of Geophysical Research: Space Physics (1978–2012)*, *110*(A3).

Zhang, n. D., Z. Xiao, K. Igarashi, and G. Ma (2002), Gps-derived ionospheric total electron content response to a solar flare that occurred on 14 july 2000, *Radio Science*, *37*(5), 19–1.

Zhang, S.-R., and J. M. Holt (2004), Ionospheric plasma temperatures during 1976–2001 over millstone hill, *Advances in Space Research*, *33*(6), 963–969.

Zhang, S.-R., J. M. Holt, A. M. Zalucha, and C. Amory-Mazaudier (2004), Midlatitude ionospheric plasma temperature climatology and empirical model based on saint santin incoherent scatter radar data from 1966 to 1987, *Journal of Geophysical Research: Space Physics (1978–2012)*, *109*(A11).

Zhou, Q. H., and M. P. Sulzer (1997), Incoherent scatter radar observations of the f-region ionosphere at arecibo during january 1993, *Journal of Atmospheric and Solar-Terrestrial Physics*, *59*(17), 2213–2229.

Zhu, J., and A. J. Ridley (2015), Investigating the performance of simplified neutral-ion collisional heating rate in a global it model, *Journal of Geophysical Research: Space Physics*, doi:10.1002/2015JA021637.

Zhu, X., E. Talaat, J. Baker, and J.-H. Yee (2005), A self-consistent derivation of ion drag and joule heating for atmospheric dynamics in the thermosphere, in *Annales Geophysicae*, vol. 23, pp. 3313–3322.

Zou, L., H. Rishbeth, I. Müller-Wodarg, A. Aylward, G. Millward, T. Fuller-Rowell, D. Idenden, and R. Moffett (2000), Annual and semiannual variations in the ionospheric f2-layer. i. modelling, in *Annales Geophysicae*, vol. 18, pp. 927–944, Springer.